

GenCore version 4.5
copyright (c) 1993 - 2000 Compugen Ltd.

Om nucleic - nucleic search, using sw model

Run on: August 6, 2002, 08:22:25 ; Search time 1704.03 Seconds
(without alignments)
3136.561 Million cell updates/sec

Title: US-09-684-215A-3
Perfect score:
Sequence: 1 acggcgccgtcgatactt.....tgcccgaggacccggcc 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*

2: em_estbum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hcc:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pth:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	65.2	16.5	289	U82114 U82114 ordered cosmid library Mycobacterium leprae genomic clone
2	55.6	14.0	603	A293428 BJ_Ba000
3	50.4	12.7	925	CNS0091P
c	49.6	12.5	12	CNS0091P
c	49.2	12.4	561	BH020987
c	48.6	12.3	634	BH361689
c	48.6	12.0	289	B864L
c	47.4	12.0	935	CNS005XK
c	47.1	11.9	1009	12 CNS001FW
c	45.2	11.4	538	BB215641
c	44.6	11.3	839	12 CNS001NB
c	44.4	11.2	504	BH190076
c	44.4	11.2	590	ATX0C60TF
c	44.4	11.2	632	BH189865
c	44.4	11.2	12	CNS0072Q
c	44.2	11.2	844	AL05652
c	44.2	11.1	493	BE355002

ALIGNMENTS

RESULT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114	U82114
DEFINITION	cosmid L-373; contig 64, DNA sequence.															
ACCESSION	U82114															
VERSION	U82114.1															
KEYWORDS	GSS.															
SOURCE	Mycobacterium leprae.															
ORGANISM	Mycobacterium leprae.															
BACTERIA	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;															
ACTINOMYCETES	Actinomycetales; Corynebacterineae; Mycobacteriaceae;															
MICROBACTERIUM	Mycobacterium.															

REFERENCE

1 (bases 1 to 289)

Silbaq,F.S., Cho,S.N., Cole,S.T. and Brennan,P.J.
Characterization of a 34-kilodalton protein of *Mycobacterium leprae* that is isologous to the immunodominant 34-kilodalton antigen of *Mycobacterium paratuberculosis*.
Infect. Immun. 66 (11), 5576-5579 (1998)

AUTHORS

JOURNAL
MEDLINE
COMMENT

Contact: Silbaq FS
Microbiology
Colorado State University
Fort Collins, CO 80523, USA

Bigmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol. Microbiol.* 7 (2), 197-206 (1993)

FEATURES

source
Class: unknown

1. location/Qualifiers
/organism="Mycobacterium leprae"
/db_xref="Taxon:1769"
/clone="cosmid L-373"; contig 64"
/clone.lib="ordered cosmid library"

BASE COUNT

62 a 83 c 93 g 51 t

ORIGIN

Query Match Similarity 16.5%; Score 65.2; DB 12; Length 289;
Best Local Similarity 55.2%; Pred. No. 0.00042;

Matches	127:	Conservative	0;	Mismatches	103;	Indels	0;	Gaps	0;	Db	483	AATCGGTCAAGGACGCCGCAGCTCGCCGCACATGGGGCATGGC3CCCCGCGA	542	
156	cgcacacaacggcaacggccacggacttcaaaacgtgtcgccggagctccggccggcaag	215	Qy	320	tcatctcgatggaccatggaaaaaagg	345								
33	CGACAAAGGCCACCCGGGCGCAAAGTATGAGTCGCGGCGGGGGCGCGA	92	Db	543	TCTGGAAGCTTAACGTTGCTGACAAG	568								
216	tctcgatctccacggggacgtgtatccggggatcgccatccggggatcgccatccggcc	275												
93	TCGCGCGTCAAGGGGGTCTGCTACTAGTGCGACGCCGCTGATGTCGACGCA	152												
276	cacccatggggggggggatcgatccggggatcgccatccggggatcgccatccggcc	335												
153	CGACCGGTTGTCGCTGCGGGCGGTCAAGGACCCGGTGACAAAGTGTGCTGAC	212												
336	gcaacaaatggggggggggatcgatccggggatcgccatccggggatcgccatccggcc	385												
213	TCGATGATGTTGAGTCACCTGGGAGG	262												
Db	220	TCGATGATGTTGAGTCACCTGGGAGG	262											
2 RESULT	2													
224	AZ934428													
225	BJ													
226	AZ934428													
227	Ba002108r													
228	B. japonicum													
229	603 bp													
230	DNA													
231	linear													
232	GSS													
233	24-Apr-2001													
234	ACCESSION													
235	DEFINITION													
236	LOCUS													
237	CNS0091P													
238	DEFINITION													
239	droso													
240	melanogaster													
241	genome survey													
242	sequence													
243	ACCESSION													
244	VERSION													
245	AI053013													
246	COMMENT													
247	TITLE													
248	ORGANISM													
249	Bradyrhizobium japonicum													
250	Bacteria; Proteobacteria; Bradyrhizobium													
251	Bradyrhizobium													
252	Bradyrhizobium japonicum													
253	Bradyrhizobium; alpha subdivision: Rhizobiaceae group;													
254	Bradyrhizobium													
255	Bradyrhizobium japonicum													
256	JOURNAL													
257	GENOME RES.													
258	11 (8), 1434-1440 (2001)													
259	REFERENCE													
260	AUTHORS													
261	Romkins, J.P., Wood, T.C., Stacey, M.G., Ioh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J. and Wind, R.A.													
262	COMMENT													
263	TITLE													
264	JOURNAL													
265	GENOME RES.													
266	11 (8), 1434-1440 (2001)													
267	MEDLINE													
268	COMMENT													
269	CONTACT													
270	Wing, R.A.													
271	Clemson University Genomics Institute													
272	Clemson University													
273	Jordan Hall, Clemson, SC 29634, USA													
274	TEL:	864 7288												
275	FAX:	864 656 4293												
276	EMAIL:	rwing@clemson.edu												
277	Class:	BAC ends												
278	High quality sequence stop:	553.												
279	FEATURES													
280	SOURCE													
281	1. . . 603													
282	/organism="Bradyrhizobium japonicum"													
283	/strain="USA110"													
284	/db_xref="taxon:375"													
285	/clone lib="B. japonicum BAC library"													
286	/lab_host="E. coli"													
287	/note="Vector: pindigo536; Site_1: HindIII"													
288	BASE COUNT													
289	ORIGIN													
290	109 a													
291	201 c													
292	203 g													
293	90 t													
294	511 others													
295	Query Match													
296	Best Local Similarity	14.0%		Score	55.6;	DB	12;	Length	603;					
297	Matches	54.4%		Pred.	No.	0.044;	Mismatches	94;	Indels	0;	Gaps	0;		
298	298	Conservative	0;	Mismatches	94;	Indels	0;	Gaps	0;					
299	299													
300	300													
301	301													
302	302													
303	303													
304	304													
305	305													
306	306													
307	307													
308	308													
309	309													
310	310													
311	311													
312	312													
313	313													
314	314													
315	315													
316	316													
317	317													
318	318													
319	319													
320	320													
321	321													
322	322													
323	323													
324	324													
325	325													
326	326													
327	327													
328	328													
329	329													
330	330													
331	331													
332	332													
333	333													
334	334													
335	335													
336	336													
337	337													
338	338													
339	339													
340	340													
341	341													
342	342													
343	343													
344	344													
345	345													
346	346													
347	347													
348	348													
349	349													
350	350													
351	351													
352	352													
353	353													
354	354													
355	355													
356	356													
357	357													
358	358													
359	359													
360	360													
361	361													
362	362													
363	363													
364	364													
365	365													
366	366													
367	367													
368	368													
369	369													
370	370													
371	371													

QY 292 ggccttaacggcatatccggtgactctcggtgacactggcaaaacaaatcgccg 351
Db 793 GTSSSDSTNSCCSACVNCCTGCGTCCSYBMCYTSRGCGGSSGGKCGGGGSS 852
QY 352 ggcacgatcaggaaacgtgacatggcgagggaccggccggc 396
Db 853 TNGMBGTSSACSSSSSSSSSSSSSSSSSSSSSSSSSSSSGV 897

RESULT 4
CNS001P/C
LOCUS CNS0091P 925 bp DNA Linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BAC19B16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4394461
KEYWORDS GSS
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
^{- Web : www.genoscope.cns.fr})
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamsoor in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RP1-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2, cn bw sp, the same strain used for the BDGP's
P1 and BSR libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center, can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES source
1. -925
/organism="Drosophila melanogaster"
/ab_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BAC19B16"
/note="end : RP13"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 12.5%; Score 49.6; DB 12; Length 925;
Best Local Similarity 12.8%; Pred. No. 0; 79;
Matches 43; Conservative 157; Mismatches 136; Indels 0; Gaps 0;

QY 60 cgggcaggcatggcatcggggatccggatccggatccggatccggatccggatcc 119
Db 900 SNSSB0SSCSBSSS1SSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 841
QY 120 ttcggggcttccgggtttccgggtgggtttcgacaaacacggacacggc 179
Db 840 BCCMCSSSSCCGSASARGYKVRAKGAKRGGGAGASHSSSACBSSSSASCW 781
QY 180 agtccacacgtggatggggatccggatccggatccggatccggatccggatcc 239
Db 780 SASSSSASSRSRSGGGAGGSSGASASSSSSSASAGSVVSSAASSSSCSCSS 721
QY 240 gatcacccggatccggatccgtggatccggatccgtggatccggatccggatcc 299

RESULT 5
RH020987 BH020987 561 bp DNA linear GSS 25-MAY-2001
LOCUS P864c_d_LEISHPAC-left.1 Leishmania major Friedlin Cosmid Genomic
DEFINITION Library Leishmania major genomic clone P864c, DNA sequence.
ACCESSION BH020987
VERSION BH020987.1 GI:14202102
KEYWORDS GSS
SOURCE ORGANISM Leishmania major.
Leishmania major.
Leishmania major.
Leishmania major.
Leishmania major Friedlin Cosmid End Sequences
COMMENT Other GSS: P864c_d.LEISHPAC-right.1
Myles, P.J., Vogt, C., Cawthon, J., Klacking, M., Marty, A., Mack, J.,
Munden, H., Nguyen, D., Robertson, L., Sisk, E., Fazelinia, G., Aggarwal,
G., Nelson, S., Seydel, A., Worthey, E. and Stuart, R.
Unpublished (2000)
Other GSS: P864c_d.LEISHPAC-right.1
Contact: Myles PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 224-8846
Fax: 206 224-0313
Email: mylespj@ebri.org
Seq primer: LEISHPAC-left
Class: PAC end
FEATURES source
1. -561
/Location/Qualifiers
/organism="Leishmania major"
/clone="P864c"
/clone.lib="Leishmania major Friedlin"
/library"
/ab_xref="taxon:5664"
/note="Vector: cIHYG Site:1: BamHI; Genomic DNA from
Leishmania major Friedlin was partially digested with
Sau3AI, size selected, and ligated with BamHI-digested
cIHYG cosmid vector DNA. 9216 clones were picked and
arrayed. Library construction is described in Evans et
al., Genomics Research, 8:135-145 (1998). The cIHYG
vector (Acc. No. CYU9231) is described in Ryan et al.,
Gene, 131:145-150 (1993)"
BASE COUNT 108 a 173 c 182 g 98 t
ORIGIN

Query Match 12.4%; Score 49.2; DB 12; length 561;
Best Local Similarity 63.6%; Pred. No. 0; 9;
Matches 75; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 159 caaacacggcaaaaggccggatccggatccggatccggatccggatccggatcc 218
Db 13 CGAGATCGGCATCCCGGGCGCTGGTACGTCGGGGCGCTGGCGCCGCGCC 72
QY 219 cggatctccatccggatccggatccggatccggatccggatccggatccggatcc 276
Db 73 CGGCATTCACCCGGCGAGCTGGTACCTGGCTAACCGCACGCCGATCAAGTCAGTC 130

RESULT 6

ACCESSION	BE361689/c	DEFINITION	BE361689 mRNA linear EST 20-JUL-2000
VERSION	DG1_81_F06_91_A002	JOURNAL	P864L
KEYWORDS	Dark Grown 1 (DG1), Sorghum bicolor cDNA, mRNA sequence.	LOCUS	P864L
AUTHORS	VERSTON EST.	DEFINITION	Leishmania major Friedlin P864, left end-sequence, genomic survey sequence.
TITLE	sorghum.	ACCESSION	AL390548
JOURNAL	Sorghum bicolor	VERSION	AL390548.1
COMMENT	Eukaryota; Viridiplantae; Streptophytia; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.	KEYWORDS	GSS.
REFERENCE	1 (bases 1 to 634)	ORGANISM	Leishmania major.
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.	JOURNAL	Leishmania major.
TITLE	An EST database from Sorghum: dark-grown seedlings	MEDLINE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
JOURNAL	Unpublished (2000)	REFERENCE	Leishmania.
COMMENT	Contact: Cordonnier-Pratt MM	AUTHORS	Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
DEPARTMENT	Department of Botany	TITLE	A physical map of the Leishmania major Friedlin genome
THE UNIVERSITY OF GEORGIA	The University of Georgia	JOURNAL	Genome Res. 8 (2), 135-145 (1998)
PLANT SCIENCES BUILDING, RM. 2502, ATHENS, GA 30602-7271, USA	Fax: 706 542 1805	MEDLINE	9816435
TEL: 706 542 1860	Email: impratteua.edu	REFERENCE	2 (bases 1 to 289)
SEQ PRIMER: PolyTMix	Sequences have been trimmed to exclude POLYA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.	AUTHORS	Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
HIGH QUALITY SEQUENCE START: 67	Seq primer: PolyTMix	TITLE	Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrel@sanger.ac.uk and alicat@sanger.ac.uk
HIGH QUALITY SEQUENCE STOP: 634	HIGH QUALITY SEQUENCE LENGTH: 198	JOURNAL	see http://www.ebi.ac.uk/parasites/leish.html
POLYA=No.	Clones to be sequenced were prepared by mass excision."	COMMENT	Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/Lmajor/
FEATURES	Location/Qualifiers	LOCATION	the primer sequence can be obtained from alicat@sanger.ac.uk.
SOURCE	1. .634 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_1="dark Grown 1 (DG1)" /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XbaI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda ZAP II. Clones to be sequenced were prepared by mass excision."	FEATURES	Location/Qualifiers
BASE COUNT	122 a 198 c 212 g 102 t	SOURCE	1. .289 /organism="Leishmania major" /strain="Friedlin" /db_xref="taxon:5664" /clone="PAC P864"
ORIGIN		BASE COUNT	49 a 86 c 108 g 46 t
FEATURES	Query Match	BASE COUNT	49 a 86 c 108 g 46 t
SOURCE	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	ORIGIN	49 a 86 c 108 g 46 t
BASE COUNT	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	RESULT	8
ORIGIN	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	LOCUS	CNS006XK
FEATURES	Query Match	LOCUS	CNS006XK
SOURCE	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR4N09 of RCI-98 library from Drosophila melanogaster (fruit fly) genomic survey sequence.
BASE COUNT	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	ACCESSION	AL060501
ORIGIN	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	VERSION	AL060501.1
FEATURES	Query Match	VERSION	GI:4945019
SOURCE	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	KEYWORDS	GSS.
BASE COUNT	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	ORGANISM	Drosophila melanogaster
ORIGIN	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterostoma; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.
FEATURES	Query Match	REFERENCE	1 (bases 1 to 935)
SOURCE	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	AUTHORS	Genoscope
BASE COUNT	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	TITLE	Direct Submission (02-JUN-1999)
ORIGIN	12.3%; Score 48.6; DB 10; Length 634; Best Local Similarity 47.0%; Pred. No. 1.2; Mismatches 150; Conservative 0; Mismatches 169; Indels 0; Gaps 0; Matches 150;	JOURNAL	Genoscope - Centre National de Sequencage : Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : RP 191 9106 ERY codex - FRANCE (E-mail : seqref@genoscope.cnrs.fr)

Tue Aug 6 11:59:24 2002

us-09-684-215a-3.rst

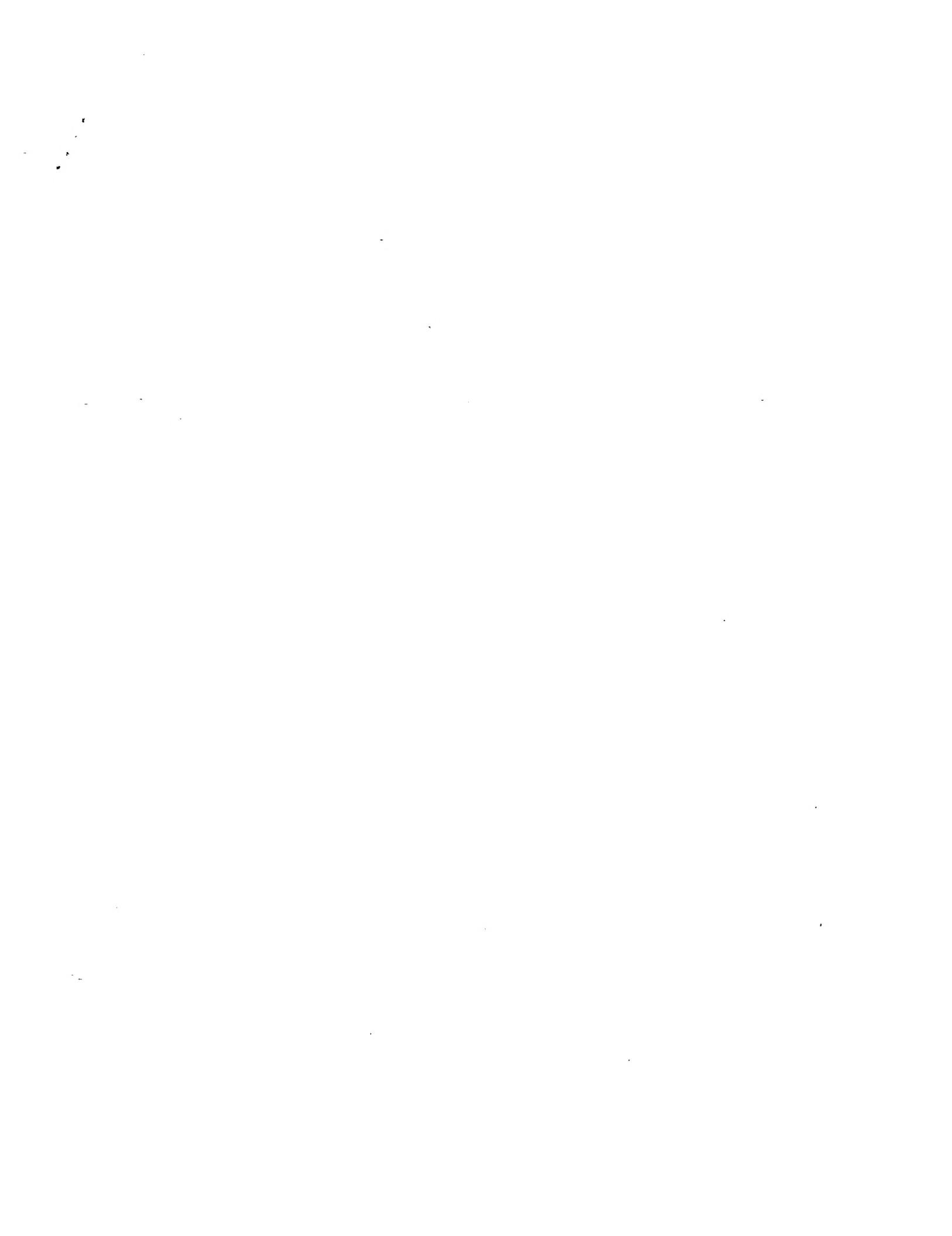
Page 9



Page 6

Tue Aug 6 11:59:21 2002

us-09-684-215a-3.rni



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 08:56:55 ; Search time 1841.62 Seconds
(without alignments)
4499.793 Million cell updates/sec

Title: US-09-684-215A-3

Perfect score: 396

Sequence: acggcccggtccgataacct.....tgcccgaggaccggcc 396

Scoring table: IDENTITY.NUC
Gapov 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Genbank:
1: gb_bat:
2: gb_htg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_Pat:
7: gb_ph:
8: gb_pi:
9: gb_pr:
10: gb_rn:
11: gb_ss:
12: gb_sy:
13: gb_un:
14: gb_vl:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mm:
20: em_on:
21: em_or:
22: em_ov:
23: em_Pat:
24: em_ph:
25: em_pl:
26: em_rn:
27: em_ss:
28: em_un:
29: em_vl:
30: em_htg_hum:
31: em_htg_inv:
32: em_htg_other:
33:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB	ID	Description
1	396	AUG05788	1068	6	AX005788	

RESULT 1
AX005788 AX005788 Sequence 907 from Patent WO9909186. DNA linear PAT 24-AUG-2000
LOCUS Sequence ACCESSION VERSION
DEFINITION AX005788 AX005788.1 GI:9928795
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium: Mycobacterium tuberculosis complex;
Mycobacterium: Mycobacterium tuberculosis
(bases 1 to 1068)
REFERENCE Portnoi,D. and Guigueno,A.
AUTHORS
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors comprising same and uses for diagnosing and preventing tuberculosis
PATENT: WO 9909186-A 97-25-FEB-1999;
PORNTOI DENIS (FR); GUIGUE NO AGNES (FR)
JOURNAL Location/Qualifiers
FEATURES source
CDS
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
1..1068
1..1068

MEDLINE
 REFERENCE
 2 (bases 1 to 11700)
 AUTHORS
 Parkhill,J.
 TITLE
 Direct submission
 JOURNAL
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique
 Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
 75124 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 COMMENT
 On 9 Jun 20, 1998 this sequence version replaced gi:2181960.
 Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URI: http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes
 implemented in tparse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.

FEATURES

Source

misc_feature

1. .11700
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:63332"

<1. .1409
 /note="fragment designated v031. Does not represent a
 physical clone"

83. .88
 /note="possible RBS, AAGGAG, for Rv0119"

gene

1. .1674
 /gene="fadD7"
 97. .1674
 /gene="fadD7"
 /note="Rv0119, (MTCI418B.01), len: 525 fadD7,
 Possible coenzyme A-ligase similar to 4'-coumarate:CoA
 ligase of many organisms. Fasta
 PS00455 putative AMP-binding domain signature. FASTA
 scores: gp|U39405|PTU39405_1|Pinus taeda xylem
 4'-coumarate:CoA (537 aa) opt: 4832; score: 526.1 E();
 8.3e-22; 28.2% identity in 440 aa overlap score is 0.896"
 /codon_start=1
 /transl_table=1
 /product="fadD7"
 /protein_id="CAB0455_1"
 /db_xref="GI:3242255"
 /db_xref="SPREMBL:007169"

RBS

gene

1. .11700
 /translation="MASDFGPRITADLVEVAATRLPEAPALVWTFADRIATISHRDRLARY
 DELAQQLURSLGLPGRDVALRMSNAEIVVALLAASRDLVWVPLDIALPTEQRVIS
 QAKKAVVLDADPHDRADPHTWPLTUVNGDSDGGTSVSHDRATEPNTAPS
 TPEGSRPDIQAMTMTGGTGTGPKVNPWHTNASSRATITGRLSPDATAVMPY
 ATETSGRKPAHLRTRCSAQSLPILPAQAGAVEIWLRGTVVRYGIDPTTQ
 AANITDTGWLTGIGLSLRAAGDLSIRGRKLINRGRKISPERVEGYLASHNMEA
 AVEGVPHOLYGDAVAVIVPRESPRPITELVWFCRELAETLIPRSPQEASGLPHTA
 KGSLURRAYAERFGHSV"

misc_feature

610. .645
 /gene="fadD7"
 /note="PS00455 Putative AMP-binding domain signature, info
 [LIMRY, {2}, STG][STG][GST][STEI][SG]. [PASLIVM][KR], info
 count = 22.0"

source

1309. .11293
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="1418B"

CDS

gene

complement(1675..3819)
 /gene="fusa2"
 complement(1675..3819)
 CDS

/gene="fusa2"
 /note="Rv0120c, (MTCI418B.02c), fusa2, len: 714 P02996
 Elongation factor g, similar to e.g. Eng_Ecoli P02996
 elongation factor g (ef-g). Escherichia coli (703 aa),
 fasta scores: opt:1049 z-score: 1105 5 E(); 0, 32.5%
 identity in 717 aa overlap. Contains P50017
 Amp/GTP-binding site motif A (P-loop) Also similar to M.
 tuberculosis fusa, MTC210_01 (39.1% identity in 299 aa
 overlap). Similar to SpfP30761 EFG_MYCLE ELONGATION FACTOR
 G (EF-G). (701 aa); 31.7% identity in 710 aa overlap."

/codon_start=1
 /transl_table=11
 /product="fusa2"
 /protein_id="CAB0448_1"
 /db_xref="GI:2181962"
 /db_xref="SWISS-PROT:Q07170"
 /translation="MADRVNSOGAAAFTPPANGPGGVNVWPGSGGKTILEA
 VAKVVISPRGSVTEGTTVCDDEAEIIRQSVGLAVASLAYDGKVNVLDPSPADFV
 GEIAGRAGAADCACVTAAGNEVDEPTSKSIAQSCQVGPMRVAWVTKLHDARANYREA
 LTAADOPAFGDKWTLPLVPLSCDGLPQESDVTPEASDVTPEASDVTPEASRGA
 LLEGIEESEDESLSMERLYLGGEEIDESVLIQDLEAVARSFSFFVPIPVSSTGNGTLE
 LEVATRGPSPMHPHLPEVTPGQHACNDAPLAVWKTSPYVAVSVL
 RVFSGTRPDITYVHGFHSSFEFGGTGSHNTHPDRDEIGVLSRPLGQQRARAVV
 AGDCAIGLUSRFEGDTSKKAELPLVWVPLPLPAAHAKTEDKLSVIGL
 KLAEDPTLRLTREIQNOHTHOVWCMGCEAHAGVWLTANLYGVSVDTFLVRLVRETF
 AGNAKGHGHRHKQSCHGQGCVCDIEVEPLPEEGSOSFERFLDKVVOGAVPROPIPVKEG
 VFAOMKGVHAGYQPVDIRVLLDKAHSVDSIDAFOAMAGALAREAAATKVILLE
 PDELSVLFDPDFGAVLDSRSRGVLTETAGHDRVVKREVPOVLETRKADLR
 SLAHGAGAFTNSAFTSPMEPESSAAVKGAG"

misc_feature

1. .11700
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:63332"

<1. .1409
 /note="fragment designated v031. Does not represent a
 physical clone"

83. .88
 /note="possible RBS, AAGGAG, for Rv0119"

gene

1. .1674
 /gene="fadD7"
 97. .1674
 /gene="fadD7"
 /note="Rv0119, (MTCI418B.03c), unknown, len: 144 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv0121c"
 /protein_id="CAB04949_1"
 /complement(3956..4390)
 /gene="fadA2"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 complement(3709..3732)
 /gene="fadA2"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 complement(3956..4390)
 /gene="fadA2"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 complement(3956..4390)
 /db_xref="Rv0121c, (MTCI418B.03c), unknown, len: 144 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv0121c"
 /protein_id="CAB04949_1"
 /complement(3956..4390)
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 complement(3956..4390)
 /db_xref="Rv0121c, (MTCI418B.03c), unknown, len: 144 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv0121c"
 /protein_id="CAB04949_1"
 /complement(4399..4401)
 /note="possible RBS upstream of Rv0121c"
 4532. .455
 /note="possible RBS upstream of Rv0122"
 /gene="Rv0122"
 4539. .4907
 /note="possible RBS upstream of Rv0122"
 /gene="Rv0122"
 4539. .4907
 /note="Rv0122, (MTCI418B.04), unknown, len: 133 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv0122"
 /protein_id="CAB0450_1"
 /db_xref="GI:2181964"
 /db_xref="SPREMBL:007172"
 /translation="MAGSVSAAAGTGWGLWVNTNRDOCYRVERTVDPALTHPEYRV
 HNGVQRMVRMRNMRKVRKRVIAVAMRRCGVPVIQEDGSLLYQGRDTSGRLTEVVA
 VRADDGDLITLHAMPKEMWR"

RBS

gene

gene

complement(1675..3819)
 /gene="Rv0122"
 /note="possible RBS upstream of Rv0123"
 4904. .5272
 /gene="Rv0123"
 4904. .5272

NETO, RAYMOND HOUGHTON, THOMAS S VEDVICK, DANIEL R TWARDZIK, PI
 PI, RAYMOND HOUGHTON, THOMAS S VEDVICK, DANIEL R TWARDZIK, PI
 MICHAEL J LODES
 PC C12N15/31, C07K14/35, C07K16/12, C12Q1/58, C12N15/62, G01N33/53 CC
 Strandness: Single;
 CC Topology: Linear;
 FH Key Location/Qualifiers
 FT source 1. .1872
 FT Location/Qualifiers
 FEATURES source
 BASE COUNT 318 a 616 c 604 g 332 t 2 others
 ORIGIN
 Query Match 98.8%; Score 391.2; DB 6; Length 1872;
 Best Local Similarity 99.2%; Pred. No. 4.2e-55; Mismatches 3; Indels 0; Gaps 0;
 Matches 393; Conservative 0;
 QY 1 acggccgcgtcgataacttcggatcgatgtggcggaggattcgccatccgatc 60
 Db 758 AGGCCCGCTGCGATACTTCAGTGCCAGGGGGCGAGATTGCGATTCGATC 817
 QY 61 ggccaggatcgatcgatcgccggatcgatcgatcgatcgatcgatcgatcgatc 60
 Db 818 GGCAAGCGAAGGGATCGGGGCCAAATCGGATCGGGGGGGTACCCACCGTCA 877
 QY 121 atcgggcttacggccatcgatcgatcgatcgatcgatcgatcgatcgatcgatc 180
 Db 878 ATGGGCCTACGCCCTCTCGCTGGTGTGACACACGGCACCGGCCACCA 937
 QY 181 gtcacacqcggtcgccggagcgatccacccggacggacatcgatcgatcgatc 240
 Db 938 GTCACACCGGTCGACGGCGCTCGATCACTCGGCACCGGAATGGGACGCCCTAAC 1057
 QY 301 ggccatcatcccgatcgatcgatcgatcgatcgatcgatcgatcgatcgatc 360
 Db 1058 GGCCATCATCCGGTGAATCTGGTGAATCTGGGACCGGT 117
 QY 361 acggggacgqgacatcgccggggaccggccggcc 396
 Db 1118 ACAGGGACGTCGACATGGCCGGGGACCCCGGCC 1153
 RESULT 10
 BD006458 LOCUS BD006458 DNA linear PAT 31-JAN-2002
 DEFINITION Compounds and methods for immunotherapy and diagnosis of Tuberculosis.

ACCESSION BD006458
 VERSION BD006458-1 GI:18634829
 KEYWORDS JPR 200101832-A/17.
 SOURCE unidentified
 ORGANISM unclassified

REFERENCE 1 (bases 1 to 1872),
 AUTHORS Reed, S.G., Skeiky, Y.A.W., Dillon, D.C., Neto, A.C., Houghton, R.,
 Tewdwick, T.S., Twardzik, D.R. and Lodes, M.J.
 TITLE Compounds and methods for immunotherapy and diagnosis of
 JOURNAL Patent: JP 200101832-A 17 13-FEB-2001;
 CORIXA CORP
 COMMENT OS Unidentified
 PN JP 200101832-A/17
 PD 13-FEB-2001
 PR 07-OCT-1997 JP 1998518456
 PR 11-OCT-1996 US 08/730510-13-MAR-1997 US 08/81812 PT
 STEVEN G REED, YASIR A W SKEIKY, DAVIN C DILLION, PI ANTONIO CAMPOS

NETO, RAYMOND HOUGHTON, THOMAS S VEDVICK, DANIEL R TWARDZIK, PI
 PI, RAYMOND HOUGHTON, THOMAS S VEDVICK, DANIEL R TWARDZIK, PI
 MICHAEL J LODES
 PC C12N15/31, C07K14/35, A61K39/00, A61K49/00, C12N15/62,
 Strandness: Single;
 CC Topology: Linear;
 FH Key Location/Qualifiers
 FT source 1. .1872
 FT Location/Qualifiers
 FEATURES source
 BASE COUNT 318 a 616 c 604 g 332 t 2 others
 ORIGIN
 Query Match 98.8%; Score 391.2; DB 6; Length 1872;
 Best Local Similarity 99.2%; Pred. No. 4.2e-55; Mismatches 3; Indels 0; Gaps 0;
 Matches 393; Conservative 0;
 QY 1 acggccgcgtcgataacttcggatcgatgtggcggaggattcgccatccgatc 60
 Db 758 AGGCCCGCTGCGATACTTCAGTGCCAGGGGGCGAGATTGCGATTCGATC 817
 QY 61 ggccaggatcgatcgatcgccggatcgatcgatcgatcgatcgatcgatc 60
 Db 818 GGCAAGCGAAGGGATCGGGGCCAAATCGGATCGGGGGGGTACCCACCGTCA 877
 QY 121 atcgggcttacggccatcgatcgatcgatcgatcgatcgatcgatcgatc 180
 Db 878 ATGGGCCTACGCCCTCTCGCTGGTGTGACACACGGCACCGGCCACCA 937
 QY 181 gtcacacqcggtcgccggagcgatccacccggacggacatcgatcgatc 240
 Db 938 GTCACACCGGTCGACGGCGCTCGATCACTCGGCACCGGAATGGGACGCCCTAAC 1057
 QY 301 ggccatcatcccgatcgatcgatcgatcgatcgatcgatcgatcgatc 360
 Db 1058 GGCCATCATCCGGTGAATCTGGTGAATCTGGGACCGGT 117
 QY 361 acggggacgqgacatcgccggggaccggccggcc 396
 Db 1058 GGCCATCATCCGGTGAATCTGGTGAATCTGGGACCGGT 117
 QY 361 acggggacgqgacatcgccggggaccggccggcc 396
 Db 1118 ACAGGGACGTCGACATGGCCGGGGACCCCGGCC 1153
 RESULT 11
 AX201049 LOCUS AX201049 DNA linear PAT 29-AUG-2001
 DEFINITION Sequence 679 from Patent WO0151633.
 ACCESSION AX201049
 VERSION AX201049.1 GI:15390857
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammal; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 675)
 AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlock, S.L., Jiang, Y.,
 Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
 Stolk, J.A., Skeiky, Y.A., Wang, A. and Meagher, M.J.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0151633-A 679 19-JUL-2001;
 CORIXA CORPORATION (US)

Tue Aug 6 11:59:19 2002

us-09-684-215a-3.rge



Query Match 99.6%; Score 394.4; DB 19; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 241 atcacccgcgtcgacggcgtccggcgcaacttcggccacccggatggcgacggcttaac 300
 ||||| 251 atcacccgcgtcgacggcgtccggcgcaacttcggccacccggatggcgacggcttaac 310
 QY 301 99gcataatccggtagtcatctggtagccaaacaaatggcgccacggat 360
 ||||| 311 99gcataatccggtagtcatctggtagccaaacaaatggcgccacggat 370

QY 361 acaggaaacctgacattggccggggaccggcc 396
 ||||| 371 acaggaaacctgacattggccggggaccggcc 406

RESULT 10
 AAV442 ID AAV442 standard; DNA; 447 BP.
 XX AC AAV442;
 XX DT 09-NOV-1998 (first entry)
 XX DE Mycobacterium tuberculosis antigen Tbra12 DNA.
 XX KW Tuberculosis; infection; diagnosis; antigen; Tbra12; ss.
 OS Mycobacterium tuberculosis strain H37Ra.
 XX FH Key Location/Qualifiers
 FT 11.406
 FT PN W098106645-A2.
 XX PD 23-APR-1998.
 XX PF 07-OCT-1997; 97WO-US18214.
 PR 13-MAR-1997; 97US 081811.
 PR 11-OCT-1996; 96US-0729622.
 PA (CORT-) CORIXA CORP.
 PT Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX DR WPI; 1998-251292/2.
 DR P-PSDB; AAW64294.

RESULT 11
 AAZ19252 ID AAZ19252 standard; DNA; 447 BP.
 XX AC AAZ19252;
 XX DT 05-NOV-1999 (first entry)
 DE M. tuberculosis antigen Tbra12 DNA sequence.
 XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test; ss.
 OS Mycobacterium tuberculosis.
 XX PN W09942076-A2.
 XX PD 26-AUG-1999.
 XX PF 17-FEB-1999; 99WO-US03268.
 PR 05-MAY-1998; 9805-0072967.
 PR 18-FEB-1998; 9805-0025197.
 PA (CORT-) CORIXA CORP.
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 DR WPI; 1999-527409/44.
 P-PSDB; AAY39096.

PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 XX skin tests and protective or therapeutic vaccines or compositions
 PS Claim 3; Page 73; 299PP; English.
 CC The present invention describes polypeptides comprising an immunogenic

This DNA sequence codes for Mycobacterium tuberculosis soluble
 CC antigen Tbra12 (see AAW64294). It was isolated from a M. tuberculosis
 CC strain H37Ra expression library with rabbit anti-sera raised
 CC against M. tuberculosis supernatant. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of an
 CC a soluble M. tuberculosis antigen, or an immunogenic portion of an
 CC M. tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic
 CC kits for detecting M. tuberculosis infection in a patient using
 CC these polypeptides, antibodies or oligonucleotide probes and
 CC primers, for the diagnosis of tuberculosis.
 XX Sequence 447 BP; 79 A; 145 C; 149 G; 72 T; 1 other;
 SQ

CC part of a *Mycobacterium tuberculosis* antigen (Ag). Also described
 CC are vaccines and fusion protein containing *M. tuberculosis* Ag's.
 CC *M. tuberculosis* Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC *M. tuberculosis* and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis immune subjects. AAZ19450 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 20; Length 447;

Best Local Similarity 99.7%; Pred. No. 4.7e-85; Mismatches 0; Indels 0; Gaps 0;

Db 1 acggcccgctcgataacttcacgtgtcccaaggatgggattccgcattccgc 60

Qy 61 gggcaggagggcgatggcgatccggatggggatccgcattccgc 70

Db 71 gggcaggatggcgatccggatacttcacgtgtcccaaggatgggattccgcattccgc 120

Qy 121 acggccgtacggcttcctggatgggttgcggatggggatccgcattccgc 180

Db 131 atccggctacccgtttctcggttgcggatggggatccgcattccgc 190

Qy 181 gtccaaacgggtggcgaggatccggatggggatccgcattccgc 240

Db 191 gtccaaacgggtggcgaggatccggatggggatccgcattccgc 250

Qy 241 atccacggcgatggcgatccggatcaacttcacgtggggatccgcattccgc 300

Db 251 atccacggcgatggcgatccggatcaacttcacgtggggatccgcattccgc 310

Qy 301 ggccatcatcccgtaacgttgcggatccggatggggatccgcattccgc 360

Db 311 ggccatcatcccgtaacgttgcggatccggatggggatccgcattccgc 370

Qy 361 acggaaacgttgcggatggggatccgcattccgc 395

Db 371 acggaaacgttgcggatggggatccgcattccgc 406

RESULT 12
AAZ1940
ID AAZ1940 standard; DNA; 447 BP.

AC AAZ1940;
DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen DNA encoding TbRa12.

XX Antigen; detection; infection; antibody; immunisation;
 KW vaccine; immunity; ss.

OS Mycobacterium tuberculosis.

PN W09942118-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-007256.

XX PR 18-FEB-1998; 98US-0024753.

XX PA (CORG-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeley YAN, Twardzik DR, Vedvick TS;
 XX WPI: 999-52741644.

DR DR P-PSDB, AAI38955.

PT PT New polypeptide comprising antigenic portions of *M. tuberculosis*

PS PS Claim 3; Page 103; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from *Mycobacterium tuberculosis*. The novel
 CC polypeptides are useful for detecting *M. tuberculosis* infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC *M. tuberculosis* infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.

XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 20; Length 447;

Best Local Similarity 99.7%; Pred. No. 4.7e-85; Mismatches 0; Indels 0; Gaps 0;

Db 1 acggcccgctcgataacttcacgtgtcccaaggatgggattccgcattccgc 60

Qy 61 gggcaggatggcgatccggatggggatccgcattccgc 70

Db 71 gggcaggatggcgatccggatacttcacgtgtcccaaggatgggattccgcattccgc 120

Qy 121 acggccgtacggcttcctggatgggttgcggatggggatccgcattccgc 180

Db 131 atccggctacccgtttctcggttgcggatggggatccgcattccgc 190

Qy 181 gtccaaacgggtggcgaggatccggatggggatccgcattccgc 240

Db 191 gtccaaacgggtggcgaggatccggatggggatccgcattccgc 250

Qy 241 atccacggcgatggcgatccggatcaacttcacgtggggatccgcattccgc 300

Db 251 atccacggcgatggcgatccggatcaacttcacgtggggatccgcattccgc 310

Qy 301 ggccatcatcccgtaacgttgcggatccggatggggatccgcattccgc 360

Db 311 ggccatcatcccgtaacgttgcggatccggatggggatccgcattccgc 370

Qy 361 acggaaacgttgcggatggggatccgcattccgc 395

Db 371 acggaaacgttgcggatggggatccgcattccgc 406

RESULT 13
AAZ1940
ID AAZ1940 standard; DNA; 447 BP.

AC AAS03780;
XX DT 28-AUG-2001 (first entry)

DE M. tuberculosis DNA encoding a partial antigen TbRa12.

XX KW TbRa12; antigen; vaccine; tuberculosis; AIDS;

XX KW acquired immunodeficiency disease; ss.

OS Mycobacterium tuberculosis.

PN

XX PD

XX PF

PR

XX PR

XX PA

Location/Qualifiers

11..409

/*tag= a

/product= "TbRa12"

PT /partial
 PT /note= "No start codon"
 XX
 XX WO200124820-A1.
 PN 12-APR-2001.
 XX DT 29-AUG-2001 (first entry)
 XX DE M. tuberculosis DNA encoding antigen HTCC#1 fusion protein #4.
 XX TBRA12-HTCC#1; antigen; vaccine; tuberculosis;
 KW AIDS; acquired immunodeficiency disease; His tag; ds.
 PR 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 OS Mycobacterium tuberculosis.
 OS Synthetic.

PA (CORI-) CORIXA CORP.
 XX
 PT Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX DR WPI; 2001-290576/30.
 DR P-ISDB; AAU01889.

XX Vaccinating against Mycobacteria infections in mammals using fusion proteins comprising combinations of heterologous antigens

XX Example 2; Page 162; 168pp; English.

XX The sequence encodes *Mycobacterium tuberculosis* TBRA12, an *M. tuberculosis* antigen. Compositions comprising at least 2 heterologous antigens, as a fusion protein, and vectors expressing the fusion proteins are used as vaccines to prophylactically immunise mammals (especially humans) against infection by mycobacteria. The compositions contain at least 2 heterologous antigens that increase the serological sensitivity of individuals infected with tuberculosis, a disease frequently affecting patients with acquired immunodeficiency disease, AIDS.

XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 22; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggccgcgtcgatcaacttccagctgtcccaaggggggcaggatcgccattccgatc 60
 CC 11 acggccgcgtcgatcaacttccagctgtcccaaggggggcaggatcgccattccgatc 70

Db 61 ggccaggatggatcgatcgccatccgatcgatcgatcgatcgatcgatcgatcgatc 120
 QY 71 ggccaggatggatcgatcgccatccgatcgatcgatcgatcgatcgatcgatcgatc 130

Db 121 atcgccgttccatcgccatccgttcgtatggatggatggatggatggatggatggatc 180
 QY 131 atcgccgttccatcgccatccgttcgtatggatggatggatggatggatggatggatc 190

QY 181 gtccaaacgtgtggagcgtccgtccgtccgtccgtccgtccgtccgtccgtccgt 240
 Db 191 gtccaaacgtgtggagcgtccgtccgtccgtccgtccgtccgtccgtccgtccgt 250

QY 241 atccacgcgttcacgggggtccgtatcaactcgccacgcgtatcgccatccgttaac 300
 Db 251 atccacgcgttcacgggggtccgtatcaactcgccacgcgtatcgccatccgttaac 310

QY 301 ggccatctccggatggatggatggatggatggatggatggatggatggatggatggatc 360
 Db 311 ggccatcatccggatggatggatggatggatggatggatggatggatggatggatggatc 370

QY 361 acggggacgtggatggatggatggatggatggatggatggatggatggatggatggatc 396
 Db 371 acaggagaactgtgacatggccaggagccccggc 406

SQ Location/Qualifiers
 FH 1..1629
 FT /*tag= a
 FT //product= "TBRA12-HTCC#1"
 FT //transl_except= "(pos:621..1623,aa:aaa)"
 FT //note= "Xaa In frame STOP codon"
 FT //partial
 FT /*note= "No start or stop codon. Although the sequence does contain an in frame stop codon, 2 further amino acids are shown in Figure 8 as being encoded by the present sequence, without a further stop codon"
 FT misc_feature 25..426
 FT /*tag= b
 FT //note= "Region derived from TBRA12"
 FT misc_feature 427..444
 FT /*tag= c
 FT misc_feature 445..1629
 FT /*tag= d
 FT //note= "Region derived from thrombin"
 FT misc_feature /*tag= d
 FT //note= "Region derived from HTCC#1"
 FT misc_feature
 XX SQ WO200124820-A1.
 PD 12-APR-2001.
 XX PF 10-OCT-2000; 2000WO-US28095.
 XX PR 07-OCT-1999; 99US-0158338.
 XX PR 07-OCT-1999; 99US-0158425.
 PA (CORI-) CORIXA CORP.
 PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX DR WPI; 2001-290576/30.

XX Vaccinating against Mycobacteria infections in mammals using fusion proteins comprising combinations of heterologous antigens

XX Example 2; Fig 8; 168pp; English.

XX The sequence encodes *Mycobacterium tuberculosis* fusion protein, TBRA12-HTCC#1 and includes a His tag at the N-terminus to aid purification. Compositions comprising at least 2 heterologous antigens, as a fusion protein, and vectors expressing the fusion proteins are used as vaccines to prophylactically immunise mammals (especially humans) against infection by mycobacteria. The compositions contain at least 2 heterologous antigens that increase the serological sensitivity of individuals infected with tuberculosis, a disease frequently affecting patients with acquired immunodeficiency disease, AIDS.

XX Sequence 1629 BP; 298 A; 500 C; 535 G; 296 T; 0 other;

Query Match 99.6%; Score 394.4; DB 22; Length 1629;
 Best Local Similarity 99.7%; Pred. No. 5.2e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggccgcgtcgatcaacttccagctgtcccaagggtggccaggatcgccatccgtatc 60

Tue Aug 6 11:59:20 2002

us-09-684-215a-3.rng



Copyright (c) 1993 - 2000 Compugen Ltd.
GenCore version 4.5

Run on: August 6, 2002, 08:08:20 ; Search time 30.6 Seconds
 Perfect score: US-09-684-215A-4
 Sequence: 675 TAASDNFQLSQGQQFAIPI. QTKSGGTRTGNVTLAEGPPA 132
 Scoring table: BLOSUM62
 GapP 10.0 , gapext: 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :

A_Geneseq_032803: *
 1: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT: *
 2: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT: *
 3: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT: *
 4: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT: *
 5: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT: *
 6: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT: *
 7: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT: *
 8: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT: *
 9: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT: *
 10: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT: *
 11: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT: *
 12: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT: *
 13: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT: *
 14: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT: *
 15: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT: *
 16: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT: *
 17: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT: *
 18: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT: *
 19: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT: *
 20: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT: *
 21: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT: *
 22: /SIDS5/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	675	100.0	231	AAV32071 Mycobacterium tube
2	675	100.0	355	Mycobacterium spec
3	675	100.0	355	Mycobacterium tube
4	675	100.0	379	Mycobacterium spec
5	675	100.0	543	M. tuberculosis an
6	675	100.0	729	Mycobacterium tube
7	670	99.3	132	Mycobacterium tube
8	670	99.3	132	Mycobacterium tube
9	670	99.3	132	Mycobacterium tube
10	670	99.3	132	M. tuberculosis im
11	670	99.3	132	Mycobacterium tube
				M. tuberculosis an

SUMMARIES

ALIGNMENTS

DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ220206.
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis -
 XX
 PS Claim 1; FIG 13A-B; 83pp; English.

This sequence represents a recombinant *Mycobacterium tuberculosis* bi-antigen fusion protein, termed Mb24, composed of the antigens Ra12 and DPPD. The fusion protein is expressed in host cells using a vector carrying a polynucleotide (see AAZ220206) comprising the coding sequences for the 2 antigens. The invention provides fusion proteins (see AAY32059-71) containing at least 2 M. tuberculosis antigens. The new fusion proteins and polynucleotides encoding them are useful as vaccines for preventing tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal skin tests for detection of anti-M. tuberculosis antibodies), monitoring of disease progression, and treatment of tuberculosis. They are more effective immunogens than mixtures of the individual protein components.

XX Sequence 231 AA;

Query Match 100.0%; Score 675; DB 20; Length 231;
 Best local Similarity 100.0%; Pred. No. 4.8e-59;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGQGFAPIQOMATAQOIQNSGGSPVTHIGPAPLGLGVWDNNGAR 60
 Db 9 taasdnhqisqggqgfaipqgamaiaqqlsqgksggsptvhigptarlgivgvvahnngar 68

QY 61 VQWVGSPASAGISGQGDPVTAQDAPINSATAMADALNGHHPDVISTWORKSGGR 120
 Db 69 vqrvgvsgspasagisgqsdgqvavdgapinsatamadalnghhpgdvistwqtksggr 128

QY 121 TGNVTLAEGPPA 132
 Db 129 tgnvvtlaegppa 140

RESULT 2
 AAY05000
 ID AAY05000 standard; Protein; 355 AA.
 XX
 AC AAY05000;
 XX
 DT 06-JUL-1999 (first entry)
 DE Mycobacterium species protein sequence 50D.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 OS Mycobacterium sp.
 WO9909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR01813.
 XX
 PR 11-SEP-1997; 97FR-0011305.
 PR 14-AUG-1997; 97FR-0010404.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
 PT Guigueno A;
 DR WPI; 1999-18045/15.
 DR N-PSDB; AAK34251.

XX
 PT Mycobacterial DNA vectors containing reporter constructs - for identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 XX
 PS Claim 32; FIG 5D; 309pp; French.

Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins from various *Mycobacterium* species microorganisms. The encoded nucleotide sequences can be used as primers and probes for methods for detecting and identifying mycobacteria, especially belonging to the *M. tuberculosis* complex. The encoded proteins can be used in vaccines for immunisation against a bacterial or viral infection.

XX Sequence 355 AA;

Query Match 100.0%; Score 675; DB 20; Length 355;
 Best local Similarity 100.0%; Pred. No. 8.3e-59;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGQGFAPIQOMATAQOIQNSGGSPVTHIGPAPLGLGVWDNNGAR 60
 Db 224 taasdnhqisqggqgfaipqgamaiaqqlsqgksggsptvhigptarlgivgvvahnngar 283

QY 61 VQWVGSPASAGISGQGDPVTAQDAPINSATAMADALNGHHPDVISTWORKSGGR 120
 Db 284 vqrvgvsgspasagisgqsdgqvavdgapinsatamadalnghhpgdvistwqtksggr 343

QY 121 TGNVTLAEGPPA 132
 Db 344 tgnvvtlaegppa 355

RESULT 3
 AAG8110
 ID AAG8110 standard; Protein; 355 AA.
 XX
 AC AAG8110;
 XX
 DT 04-SEP-2001 (first entry)
 DE Mycobacterium tuberculosis potential drug target protein SEQ ID 161.
 KW Drug target; growth; organism viability; characterisation.
 OS Mycobacterium tuberculosis.
 XX
 PN WO00135317-A1.
 XX
 PD 17-MAY-2001.
 XX
 PF 13-NOV-2000; 2000WO-US31152.
 XX
 PR 12-NOV-1999; 99US-0165006.
 PR 12-NOV-1999; 99US-0165124.
 PR 01-FEB-2000; 2000US-0179531.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Eisenberg D, Rotstein SH, Marcotte EM;
 XX
 DR WPI; 2001-329193/34.
 DR N-PSDB; AAH81961.

Identifying nucleotide or polypeptide sequence for use as drug target, involves providing algorithm that analyzes a functional relationship between nucleotide or polypeptide sequences, and comparing the sequences - Disclosure; Page 157; 207pp; English.

This invention relates to a method for identifying a nucleotide or

CC antigens, as a fusion protein, and vectors expressing the fusion
 CC proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS.

XX Sequence 543 AA;

Query Matchⁿ 100.0%; Score 675; DB 22; Length 543;
 Best local Similarity 100.0%; Pred. No. 1 4e-58; Mismatches 0; Indels 0; Gaps 0;

Matches 132; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFOLSQGCGFAPIGORMAIAQGQINSGGGSPVHIGPTAFGLGVDDNNNGAR 60
 Db 9 taasdnfqisqggqgfaipggmalaqgqrsqggspvhlqptafglgvddnnngar 68

OY 61 VQRVGSAPASLGISGTSQGDTAVDGAPINSATAMADALNHRGDSVWTKSGTR 120
 Db 69 vqrvgsgapasiqgtsqgdvavtadvgapinsatamadalingphgpdvisvtwqtksgtr 128

OY 121 TGNVTLAEGPPA 132
 Db 129 tgnvtlaegppa 140

RESULT 6

ID AAY32059
 XX AAY32059 standard; Protein; 729 AA.

AC AAY32059;
 XX

DT 17-JAN-2000 (first entry)

XX DE Mycobacterium tuberculosis antigen fusion protein Mtb32A.

XX KW Tuberculosis; antigen; fusion protein; Mtb32A; Ra12; TbH9; Ra35;
 KW diagnosis; therapy; vaccine; immunogen.

OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers

FT Peptide 1..7 /note= "Met/His tag"
 FT Protein 8..139 /note= "Ra12"

FT Protein 142..532 /note= "TbH9"
 FT Protein 535..729 /note= "Ra35"

XX PN W0951748-A2.

PD 14-OCT-1999.

XX PF 07-APR-1999; 99WO-US07717.

XX PR 07-APR-1998; 9805-0056556.
 PR 30-DEC-1998; 9805-0223040.

XX PA (CORTI-) CORIXA CORP.

XX PT Skeiky YAW, Alderson M, Campos-Neto A;

DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ20194.

XX New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis - are
 XX

PS Clair-1; Fig 1A-B; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis
 CC tri-antigen fusion protein, termed Mt32A, composed of the antigens
 CC Ra12, TbH9 and Ra35. The fusion protein is expressed in host cells
 CC using a vector carrying a polynucleotide (see AAZ20194) comprising
 CC the 3 coding sequences for the antigens. The invention provides
 CC fusion proteins (see AAY32059-71) containing at least 2 M.
 CC encoding them are useful as vaccines for preventing tuberculosis
 CC (claimed), for diagnosis (via in vitro assays or intradermal skin
 CC tests for detection of anti-M. tuberculosis antibodies), monitoring
 CC of disease progression, and treatment of tuberculosis. They are
 CC more effective immunogens than mixtures of the individual protein
 CC components.

SQ sequence 729 AA;

XX Query Match 100.0%; Score 675; DB 20; Length 729;
 Best Local Similarity 100.0%; Pred. No. 2.1e-58; Mismatches 0; Indels 0; Gaps 0;

Matches 132; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFOLSQGCGFAPIGORMAIAQGQINSGGGSPVHIGPTAFGLGVDDNNNGAR 60
 Db 8 taasonfqisqggqgfaipggmalaqgqrsqggspvhlqptafglgvddnnngar 67

OY 61 VQRVGSAPASLGISGTSQGDTAVDGAPINSATAMADALNHRGDSVWTKSGTR 120
 Db 68 vqrvgsgapasiqgtsqgdvavtadvgapinsatamadalingphgpdvisvtwqtksgtr 127

RESULT 7

ID AAW32422
 XX AAW32422 standard; Protein; 132 AA.

AC AAW32422;
 XX DT 08-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen TbRa12.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M. tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN W09709428-A2.

XX PD 13-MAR-1997.

XX PF 30-APR-1996; 96WO-US14674.

XX PR 12-JUL-1996; 96US-0680574.
 PR 01-SEP-1995; 95US-0524336.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

XX PA (CORTI-) CORIXA CORP.

XX PT Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PT Twardzik DR, Vedick TH;

XX DR WPI; 1997-19293/17.

XX PT N-PSDB; AAT91466.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevent or treatment of tuberculosis, also
 PT for diagnosis

PS Example 3; Page 96-97; 168pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M. tuberculosis
CC antigen, TbRa12. The immunogenic protein, and fusion proteins
CC containing one or more of the proteins or one of the proteins plus
ESAT-6, are useful in vaccines, preferably when formulated with a
CC non-specific adjuvant, to induce an immune response against
CC M. tuberculosis (for treatment or prevention).
SQ sequence 132 AA;

Query Match 99.3%; Score 670; DB 18; length 132;
Best Local Similarity 99.3%; **Pred.** No. 7. 4e-59;
Matches 131; **Conservative** 0; **Mismatches** 1; **Indels** 0; **Gaps**
OY 1 TAASINFOLSGCGGFAPIGOMATAQOISGGGSPVHGPATARFLGIVDNNNGAR 60
DB 1 taasinfqsgggfaipiogramataqoisgggspvhgpatarflgivdnnngar 60
OY 61 VQRVGSAPASLGISTGVDITAVDGAPINSATAMDALNHSHPGVISWQTKSGGTR 126
DB 61 vqrvgsgsapaslgistgvditavdgapinsatamadalnhpgviswqtksggr 126
OY 121 TGNVTLAEGPPA 132
DB 121 tgnvtlaegppa 132

RESULT 8
ID AAW32354
ID AAW32354 standard; Protein; 132 AA.
XX
AC AAW32354;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbRa12.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M. tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN W0970942B-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14675.
XX
PR 12-JUL-1996; 96US-0650573.
PR 01-SEP-1995; 95US-0523435.
PR 22-SEP-1995; 95US-0522136.
PR 22-MAR-1996; 96US-0650280.
PR 05-JUN-1996; 96US-0658800.
XX
PA (CORI-) CORINA CORP.
XX
PI Campes-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAH;
XX Twardzik DR, Vedvick TH;
DR WPI: 1997-19294/17.
DR N-P5DB; AAT91403;
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
PT - useful for diagnosis of M. tuberculosis infection
PS Example 3; Page 103; 190pp; English.

Query Match	Score	DB	Length
Best Local Similarity	670;	18;	132;
Matches	Pred. No.	Mismatches	Gaps
1	TAASONFQLSQGGFAPIPGOMAATGQIRSSGGSPVHIGPNAFLGLGVVDNNNGCAR	60	0;
Db	taasdnfqisqgggfaipiagammaiaqgirssggspvngprafqglgvvdnnngar	60	0;
QY	61 VORVGSGAPASIGISTGDVIAVDGAPINATAMADALINGHHPGDNISVWIKTKSGR	120	0;
Db	61 vqrvvsgapasiqgdstgdvitaqvadgapinsatamadalinghpgdvisvnqtksggr	120	0;
QY	121 TGNVTLAEGPA 132		
Db	121 tgnvilaedppa 132		
RESULT	9		
ID	AAWB1657 standard; protein: 132 AA.		
XX			
AC	AAWB1657;		
XX			
DT	27-JAN-1999 (first entry)		
DE	M. tuberculosis immunogenic polypeptide TbRa12.		
XX			
KW	Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB; vaccine; pharmaceutical; infection; diagnosis.		
XX			
OS	Mycobacterium tuberculosis.		
XX			
PN	WO9816646-A2.		
XX			
PD	23-APR-1998.		
XX			
PF	07-00T-1997; 97M0-US18293.		
XX			
PR	13-MAR-1997; 97US-0818112.		
PR	11-OCT-1996; 96US-0730510.		
XX			
PA	(CORT-) CORIXA CORP.		
XX			
PI	Campos-Neto A, ¹ Dillon DC, ¹ Houghton R, ¹ Lodes MJ, ¹		
PI	Reed SG, ² Skeley YAW, ² Twardzik DR, ² Vedvick TS, ¹		
XX			
DR	WPI: 1998-261042/23.		
DR	N-PSB; AAV64450.		
XX			
PT	Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to develop products for the detection of M. tuberculosis infection		
PT	and for diagnosis, treatment and prevention of tuberculosis		
XX			
PS	Example 3; Page 97-98; 230pp; English.		
CC	This sequence represents an immunogenic portion of a soluble Mycobacterium tuberculosis (MT) antigen which can be used in a method for inducing protective immunity against tuberculosis (TB). This sequence can be formulated into vaccines and/or pharmaceutical compositions for		

CC immunising against *M. tuberculosis* infection or may be used for the diagnosis of tuberculosis.

CC XX Sequence 132 AA;

Query Match 99.3%; Score 670; DB 19; Length 132;
Best Local Similarity 99.2%; Pred. No. 7.4e-59; 0; Mismatches 0; Indels 0; Gaps 0;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGOGFAPIGQMAIQAGQIRGGGSPTVHGPATFLGLSYVDNINGAR 60
Db 1 taasdnfqlsqgqgfaipiiggamaiagqisgggsptvnhgptafqglgvvdningar 60
QY 61 VQRTVGSAAASIGISTDVTAVDGAPINSAATAMDALNHRHPGVISVWQTSGGTR 120
Db 61 vqrvgsgapaaasigistgdvtavdgapinstatamadalnhrhpgrisvwtsgtr 120
QY 121 TGNVTLASGPAA 132
Db 121 tgnvtlaeqpaa 132

Db 121 tgnvtlaeqpaa 132

RESULT 10

AAW64294
ID AAW64294 standard; Protein; 132 AA.
AC AAW64294;
XX

DT 09-NOV-1998 (first entry)
DE Mycobacterium tuberculosis antigen Tbra12.
XX

KW Tuberculosis; infection; diagnosis; antigen; Tbra12.
XX

OS Mycobacterium tuberculosis strain H37Ra.
XX

PN W09816645-A2.
XX

PD 23-APR-1998.
XX

PF 07-OCT-1997; 97WO-US18214.
XX

PR 13-MAR-1997; 97US-081811.
11-OCT-1996; 96US-0729622.
XX

PA (CORI-) CORIXA CORP.
XX

PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
DR N-PSDB; AAV44342.
XX

PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to develop products for the detection of *M. tuberculosis* infection and diagnosis of tuberculosis
XX
Example 3; Page 101; 250pp; English.
PS XX

CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen Tbra12. It is encoded by a DNA sequence (see AAV44342) isolated from a *M. tuberculosis* strain H37Ra expression library with rabbit anti-sera raised against *M. tuberculosis* supernatant. The invention relates to compositions and methods for diagnosing tuberculosis. It provides Polypeptides (see AAW64291 WO4379) comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or an immunogenic portion of an *M. tuberculosis* antigen, as well as DNA sequences encoding such polypeptides, recombinant expression vectors and transformed or transfected host cells. Also claimed are methods and diagnostic kits for detecting *M. tuberculosis* infection in a patient using the above polypeptides, antibodies, or oligonucleotide probes and primers, for the

CC diagnosis of tuberculosis.
CC XX Sequence 132 AA;

Query Match 99.3%; Score 670; DB 19; Length 132;
Best Local Similarity 99.2%; Pred. No. 7.4e-59; 0; Mismatches 0; Indels 0; Gaps 0;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGOGFAPIGQMAIQAGQIRGGGSPTVHGPATFLGLSYVDNINGAR 60
Db 1 taasdnfqlsqgqgfaipiiggamaiagqisgggsptvnhgptafqglgvvdningar 60
QY 61 VQRTVGSAAASIGISTDVTAVDGAPINSAATAMDALNHRHPGVISVWQTSGGTR 120
Db 61 vqrvgsgapaaasigistgdvtavdgapinstatamadalnhrhpgrisvwtsgtr 120
QY 121 TGNVTLASGPAA 132
Db 121 tgnvtlaeqpaa 132

Db 121 tgnvtlaeqpaa 132

RESULT 11

AAV39096
ID AAV39096 standard; Protein; 132 AA.
AC AAV39096;
XX

DT 05-NOV-1999 (first entry)
XX

DE M. tuberculosis antigen Tbra12 amino acid sequence.
XX

KW Mycobacterium tuberculosis; *M. tuberculosis*; antigen; immunogen; immunotherapy; diagnosis; vaccination; infection; immune response; skin test.
XX

OS Mycobacterium tuberculosis.
XX

PN W09942076-A2.
XX

PD 26-AUG-1999.
XX

PF 17-FEB-1999; 99WO-US03268.
XX

PR 05-MAY-1998; 98US-0072967.
18-FEB-1998; 98US-0025197.
XX

PA (CORI-) CORIXA CORP.
XX

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R; PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
DR N-PSDB; AAZ19252.
XX

PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin tests and protective or therapeutic vaccines or compositions
XX
Example 3; Page 98; 299pp; English.
PS XX

CC The present invention describes polypeptides comprising an immunogenic part of a Mycobacterium tuberculosis antigen (Ag). Also described are vaccines and fusion protein containing *M. tuberculosis* Ag's. *M. tuberculosis* Ag's, DNAs encoding them, derived fusion proteins and other polypeptides fragments, can be used in pharmaceutical compositions or vaccines to generate a protective or therapeutic immune response to *M. tuberculosis* and as reagents in skin tests for diagnosis of tuberculosis. Ag can induce proliferation of, or cytokine secretion by, T, B or natural killer cells and/or macrophages in *tuberculosis* immune subjects. AAZ1925 to AA19460 and AAV39083 to AAV39225 are used in the exemplification of the present invention.

CC Sequence 132 AA;

Query Match 99.3%; Score 670; DB 20; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59; Mismatches 131; Conservative 0; Indels 0; Gaps 0;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOUSQGGCGFAPIGQAMAIAQOIRSGGGSPVHIGPFAFIGLGIVWDNGNGAR 60
 Db 1 taasdnfqisqggggfaipggamaeqirsgggspvhwigtaflglgvvdngar 60

QY 61 VQRVVGASAPASLGLISTGDVITAVDGAPINSATAMADAINGHHPGDVISWTQKSGTR 120
 Db 61 vqrvgasapaslglstgdvitaqvadgapinsatamadainghhpgdvisnwqtksgr 120

QY 121 TGNVTLAEGPPA 132
 Db 121 tgnvtlaegppa 132

RESULT 12
 AAY38959 standard; protein; 132 AA.
 ID AAY38959
 XX AC AAY38959;
 XX DE M. tuberculosis recombinant antigen protein Tbra12.
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 KW Mycobacterium tuberculosis.
 XX PN W09942118-A2.
 XX PD 26-AUG-1999.
 XX PF 05-MAY-1999; 99WO-US03265.
 XX PR 18-FEB-1998; 98US-0024753.
 XX PA (CORI-) CORIXA CORP.
 XX PA Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PT PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX DR WPI: 1999-52741644.
 DR N-PSDB; AAZ19040.

XX PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX PS Example 3; Page 138; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from *Mycobacterium tuberculosis*. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX Sequence 132 AA;

Query Match 99.3%; Score 670; DB 20; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59; Mismatches 131; Conservative 0; Indels 0; Gaps 0;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOUSQGGCGFAPIGQAMAIAQOIRSGGGSPVHIGPFAFIGLGIVWDNGNGAR 60
 Db 1 taasdnfqisqggggfaipggamaeqirsgggspvhwigtaflglgvvdngar 60

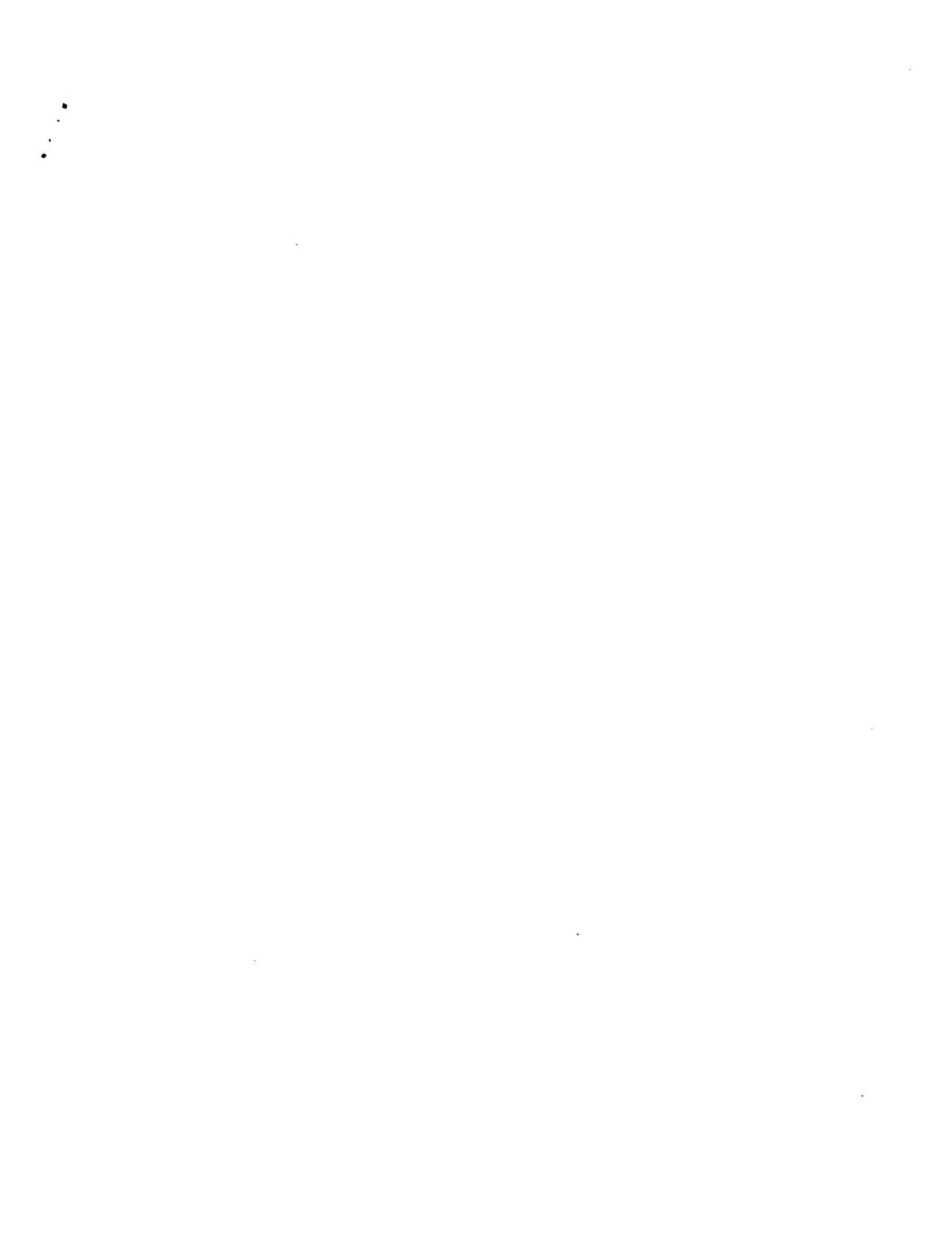
RESULT 13
 AAU69898 standard; protein; 132 AA.
 ID AAU69898
 XX AC AAU69898;
 XX DT 30-JAN-2002 (first entry)
 XX DE Mycobacterium tuberculosis antigen Ra12.
 XX KW Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;
 KW fusion protein; Ra12 antigen.
 XX OS Mycobacterium tuberculosis.
 PN WO200173032-A2.
 XX PD 04-OCT-2001.
 XX PF 27-MAR-2001; 2001WO-US09919.
 XX PR 27-MAR-2000; 2000US-0536857.
 PR 09-MAY-2000; 2000US-0568100.
 PR 12-MAY-2000; 2000US-0570737.
 PR 13-JUN-2000; 2000US-0593793.
 PR 27-JUN-2000; 2000US-0505783.
 PR 10-AUG-2000; 2000US-0535215.
 PR 29-AUG-2000; 2000US-061236.
 PR 06-SEP-2000; 2000US-0657279.
 PR 02-OCT-2000; 2000US-0679426.
 PR 10-OCT-2000; 2000US-0685166.
 XX PA (CORI-) CORIXA CORP.
 XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD,
 PT Fainger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
 XX DR WPI: 2001-63923273.
 XX PT New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer.
 XX PS Example 17; Page 531-532; 579pp; English.

XX The invention relates to isolated prostate-specific
 CC polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen-presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumour protein and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is M. tuberculosis Ra12
 CC antigen (or fragment) used to make a fusion protein comprising a
 CC prostate specific polypeptide of the invention.
 XX Sequence 132 AA;

CC NAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX

	Query Match	Best Local Similarity	Score	DB	Length
OY	TRASDNEQLSGQGFAPIPGOMATAGQIQRSGGSPVWHIGTAFLGLGVVDNNNGAR	99.3%	670	22	132;
Db	taasdlnfqslqqgqfaipqgamaiaqqrtsqggspvhqgptafqglgvadnngar	99.2%	Pred. No.	7.4e-59	
OY	61 VQRVVVASAPAMSLGISTGDVITAVDGAPINSATAMADALNHGHPGDVITSVTWQTKSGGR	120			
Db	61 vqrvvgasapaslglstegdvitavdgapinsatamadalnhghpgdvisvnwqtksggr	120			
OY	TGNVTIAEGPPA	132			
Db	tgnvtlaegppa	132			

Search completed: August 6, 2002, 08:15:50
Job time: 450 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:12:40 ; Search time 19.71 Seconds
(without alignments)
643.521 Million cell updates/sec

Title: US-09-684-215A-4
Perfect score: 675
Sequence: 1 TAASDNFQLSQGGQFAIPI.....QTKSGGTRTGNVTLAEGPPA 132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : pIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	675	100.0	355	F70983	probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37Rv)
2	474.5	70.3	361	S47170	hypothetical protein
3	459.5	68.1	354	A87242	probable secreted
4	199.5	29.6	464	C70821	probable serine proteinase
5	188.5	27.9	382	H86930	probable secreted
6	188.5	27.9	452	T45448	probable serine proteinase hhca (E. coli K12)
7	139	20.6	394	S74613	probable secreted
8	139	20.6	407	AG2150	probable secreted
9	138.5	20.5	362	T35287	probable secreted
10	133.5	19.8	408	H86891	exported serine proteinase
11	131	19.4	441	E73357	probable periplasmic proteinase
12	128.5	19.0	514	A82581	periplasmic proteinase
13	122.5	18.1	474	F83550	serine proteinase
14	119.5	17.7	203	T35866	probable integral membrane protein
15	117.5	17.4	530	F87590	serine proteinase
16	113	16.7	393	E95261	serine proteinase
17	113	16.7	397	B98127	serine proteinase
18	113	16.7	513	A00060	serine proteinase
19	113	16.7	513	AD3418	probable serine proteinase (EC 3.4.23.1)
20	113	16.7	523	A97479	probable serine proteinase
21	113	16.7	523	A12696	serine proteinase
22	112	16.6	429	AD1894	serine proteinase
23	111	16.4	455	C91142	serine endopeptidase
24	111	16.4	455	F88597	serine endopeptidase
25	111	16.4	499	B8194	probable periplasmic proteinase
26	109	16.1	455	AB0909	serine protease (E. coli K12)
27	108	16.0	348	H96526	serine protease
28	108	16.0	472	C87408	serine proteinase
29	107.5	15.9	452	S77538	serine proteinase

ALIGNMENTS

RESULT 1
F70983
C Species: Mycobacterium tuberculosis
C Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C Accession: F70983
R Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, ReJandream, M.A.; Rossers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A Reference number: A70500; MUID:98295987
A Accession: F70983
A Status: preliminary; nucleic acid sequence not shown; translation not shown
A Molecule type: DNA
A Residues: 1-355 <CDS>
A Cross-references: GB:Z96071; GB:AL123456; NID:93242254; PIDN:CA09453.1; PIB:g21819
A Experimental source: Strain H37Rv
C Genetics:
A Gene: pepA
C Superfamily: Escherichia coli trypsin-like proteinase

Query Match 100.0%; Score 675; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 9.2e-49;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSQGGQFAIPIGGQAMATAGOIRSRRGGSSPTWINGPAAFLGLGQWWDNGAR 60
Dy 224 TAASDNFQLSQGGQFAIPIGGQAMATAGOIRSRRGGSSPTWINGPAAFLGLGQWWDNGAR 283

QY 61 VORVGSAPASLIGSTGMDVITAVDGAPINSATAMADALINGHHPDDIVSTWQKSGGR 120
Dy 284 VQRVGSAPASLIGSTGMDVITAVDGAPINSATAMADALINGHHPDDIVSTWQKSGGR 343

QY 121 TCNTVLAEGPPA 132
Dy 344 TCNTVLAEGPPA 355

RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C Species: Mycobacterium paratuberculosis
C Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999
C Accession: S47170
R Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
A Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratuberculosis
A Reference number: S47170
A Accession: S47170

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-361 <CAM>
A;Cross-references: EMBL:Z23092; NID:g505550; PIDN:CAA80638.1; PID:g505551
C;Superfamily: proteinase hhOB

Query Match 70.3%; Score 474.5; DB 2; Length 361;
Best Local Similarity 69.7%; Pred. No. 3.9e-32; Mismatches 17; Indels 1; Gaps 1;

QY 1 TAASDNFQLSQGQFAPIGOMAIAGQIQRGGSPVHIGPTAFGLGVDDNNNGAR 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 TAATDSIKMS-GGQGFAPIGOMAYANQIQRGGSPVHIGPTAFGLGVDDNNNGAR 289

QY 61 VORVVGSAAPAASLGIISGQDVTAHDGAPINSATAMADALNHHPGIVSIVWQTKSGTR 120
|||||: |||: |||: |||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 280 VQVNVNGPAAANGIAINFODVITVQDVTIPINGATSMIEVLVPHPGDTIAVFRSYDGER 349
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 121 TGNVTLAEGPPA 132
|||:
Db 350 TANITLAEGPPA 361
|||:

RESULT 3

A8242 probable secreted serine proteinase [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae
C;Accession: A8242
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
R.; Davies, R.M.; Devlin, K.; Dutchoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001.

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-354 <STO>
A;Cross-references: GB:AL450380; NID:913093863; PIDN:CAC32191.1; GSPDB:GN00147
A;Gene: ML2659

Query Match 69.1%; Score 459.5; DB 2; Length 354;
Best Local Similarity 66.7%; Pred. No. 6.7e-31; Mismatches 17; Indels 26; Gaps 1;

QY 1 TAASDNFQLSQGQFAPIGOMAIAGQIQRGGSPVHIGPTAFGLGVDDNNNGAR 60
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 224 TAATDNYKM-LGGQFAPIGOMAEVVGAIRSGAGSNTRHIGPTAFGLGVDDNNNGAR 282

QY 61 VORVVGSAAPAASLGIISGQDVTAHDGAPINSATAMADALNHHPGIVSIVWQTKSGTR 120
|||||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 283 VARYVNGPAAANGIAINFODVITVQDVTIPINGATSMIEVLVPHPGDTIAVFRSYDGER 342
|||:
QY 121 TGNVTLAEGPPA 132
|||:
Db 343 TANITLAEGPPA 354
|||:

RESULT 4

C70821 probable serine proteinase Rv0993 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: C70821
R.; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Devlin, K.; Feil, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Shelton, S.; Squares, S.; Nature 393, 537-544, 1998

Query Match 29.6%; Score 199.5; DB 2; Length 464;
Best Local Similarity 38.9%; Pred. No. 3.2e-09; Mismatches 51; Indels 9; Gaps 3;

QY 1 TAASDNFQLSQGQFAPIGOMAIAGQIQRGGSPVHIGPTAFGLGVDDNNNGAR 60
|||||: |||: |||: |||: |||: |||: |||:
Db 337 TLGSDASDAGSGSGLGFAIPVDOAKRILADELISTGKA---SHASLGYQVNTDKTL 390

QY 58 GARYQVVGSAAPAASLGIISGQDVTAHDGAPINSATAMADALNHHPGIVSIVWQTKSG 117
|||: |||: |||: |||: |||: |||: |||: |||:
Db 391 GAKIVVWVGAGAANAGYQPKGVVYKVDPRPINSADALVAVRSKAGTAVLTQDPG 450
|||:
QY 118 GTRGCVNLAE 128
|||:
Db 451 GSRTVQVQTLGG 461
|||:

RESULT 5

H86930 probable secreted serine proteinase [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae
C;Accession: H86930
R.; Cole, S.T.; Elgelmier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; R.; Davies, R.M.; Devlin, K.; Dutchoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001.

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-382 <STO>
A;Cross-references: GB:AL450380; NID:913092536; PIDN:CAC29684.1; GSPDB:GN00147
A;Gene: MN0176
C;Superfamily: Escherichia coli trypsin-like proteinase

Query Match 27.9%; Score 188.5; DB 2; Length 382;
Best Local Similarity 33.8%; Pred. No. 2.1e-08; Mismatches 51; Indels 7; Gaps 3;

QY 2 AASDNFQLSQGQFAPIGOMAIAGQIQRGGSPVHIGPTAFGLGVDDNNNGAR 60
|||: |||: |||: |||: |||: |||:
Db 258 ADSGDAQSSIGLGFIAIPVDOAKRILADELISTGKA---ASLGQVADKGTPGAK 311
|||:
QY 61 VORVVGSAAPAASLGIISGQDVTAHDGAPINSATAMADALNHHPGIVSIVWQTKSGTR 120
|||: |||: |||: |||: |||: |||:
Db 312 VMDVYAGGAAANAAPVKGVVLTKVDRDLISADALVAVRKAPGDKVSLTYQDQGSSR 371
|||:
QY 121 TGNVTLAE 128
|||:
Db 372 TWQVNLGK 379
|||:

RESULT 6

T45448 probable serine proteinase (BC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C;Species: Mycobacterium leprae
 C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000
 C;Accession: T45448
 R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1998
 A;Reference number: 222967
 A;Accession: T45448
 A;Status: preliminary; translated from GB/EMBL/DDJB
 A;Molecule type: DNA
 A;Residues: 1-452 <JMA>
 A;Cross-references: EMBL:AL035500; PIDN:CAB36690.1
 A;Experimental source: cosmid L373
 C;Genetics:
 A;Note: MLCB373_28
 C;Keywords: hydrolase; serine proteinase
 F;182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 27.9%; Score 188.5; DB 2; Length 452;
 Best Local Similarity 39.8%; Pred. No. 2.6e-08;
 Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

QY 2 AASDNFQLSOGGAFPAIPQGAMIAQGQTSGGGSPVHIGPTAFLGLGVVDNNNG-GAR 60
 Db 328 ADSGAQOSGSIGLGFRAPIVDPQARRADELISTG--KATH---ASLGQVADKGTPGAK 381

QY 61 VQRVVGSPAPASLGIYSTGDVITAVDGAPINSATAMADALNHGHPIGDVSVTWTQTKSGTR 120
 Db 382 VMDVVAGGAAANAAVPKGVLKVDRLTSSADAVLAVERSKARPDKVSLTVQDQSGSSR 441

QY 121 TGNVTLAE 128
 Db 442 TVQVPLGK 449

RESULT 7

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O., K.; Okumura, S.; Shimoji, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yanada, M.; Yasuda
 DNA Res. 3, 105-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 s.

A;Reference number: S74643
 A;Accession: S74643
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-394 <RAN>
 A;Cross-references: EMBL:D90900; GB:AB001339; NTD:g1651768; PIDN:BA16795.1; PID:9165186
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Genetics:
 A;Gene: hhoA
 C;Superfamily: proteinase hhoB
 C;Keywords: hydrolase; proteinase

Query Match 20.6%; Score 139; DB 2; Length 394;
 Best Local Similarity 34.6%; Pred. No. 0.00028; Matches 45; Conservative 10; Mismatches 43; Indels 32; Gaps 5;

QY 13 GGGFAIPPIGQAMIAQGQIRSGGSPVHIGPTAFLGLGV---VNDNGN-----G 58
 Db 261 GIGGFAIPIDOKAONTLAAAGGTVPHPYIG-VQMMNTDQOQNRPNSPEIPEVGD 319

QY 59 ARVQVVGSPAPASLGIYSTGDVITAVDGAPINSATAM-----ADALNHGP 104
 Db 320 ILVMVRLPGTIPAAERAGIRRGDVTIVVNDGTPISDGAQLQRIIVEQAGLNKALKDLRGDR 379

RESULT 8

QY 105 GDVISVTVQQT 114
 Db 380 ---LSLTWOT 386

RESULT 8

QY AG2150
 C;Species: Anabaena sp.
 A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C;Accession: AG2150
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itoigawa, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-407 <KUR>
 A;Experimental source: strain PCC 7120
 A;Cross-references: GB:BA000019; PIDN:BA874457.1; PID:917131851; GSPDB:GN00179
 A;Genetics:
 A;Gene: alr2758

Query Match 20.6%; Score 139; DB 2; Length 407;
 Best Local Similarity 33.3%; Pred. No. 0.00029; Matches 45; Conservative 17; Mismatches 47; Indels 26; Gaps 5;

QY 13 GGGFAIPPIGQAMIAQGQIRSGGSPVHIGPTAFLGLGV-----VDNNG--- 56
 Db 275 GIGGFAIPIDOKAONTLQERDG--KVAH---PYLGQVQMLTPELAQQNNDNSAFAI 328

QY 57 ---NGARYVQVGSPAPASLGIYSTGDVITAVDGAPINSATAMADALNHGHPIGDVSVTWO 113
 Db 329 PEVNGVVLVIRVNPNSPANAGIRRQDVILOVQGQAITAEOLOQVNENSRQLQALQVRLQ 388

QY 114 TNSGGTRGRGNVLAE 128
 Db 389 -RGNOTQOOLSVRITA 402

RESULT 9

R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A;Reference number: Z21574

C;Accession: T35287
 C;Species: Streptomyces coelicolor
 C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
 C;Accession: T35287
 R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A;Reference number: T35287
 A;Status: preliminary; translated from GB/EMBL/DDJB
 A;Molecule type: DNA
 A;Residues: 1-362 <SEE>
 A;Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN00070; SCOEDB:SC5F7.30
 A;Experimental source: strain AJ(2)
 C;Genetics:
 A;Gene: SCOEDB:SC5F7.30
 C;Superfamily: Escherichia coli trypsin-like proteinase

Query Match 20.5%; Score 138.5; DB 2; Length 362;
 Best Local Similarity 35.5%; Pred. No. 0.00028; Matches 43; Conservative 14; Mismatches 55; Indels 9; Gaps 4;

QY 13 GGGFAIPPIGQAMIAQGQIRSGGSPVHIGPTAFLGLGV-----GARVQVVG 67
 Db 245 GIGGFAIPASMVTVAGQVLRDCKVTSGRAGA---RTVWDDSYRPAGAVVEVSDG 301

QY 68 APASLGLGISTGDVITAVDGAPINSATAMADALNHGHPIGDVSVTWTQTKSGTRGNVLA 127

Db 302 GADDAGLRLPGDWLVKLGDTDTITINSLSEALSMRPGDRTRKVY-TRDGKEHTAETVLG 360
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 128 E 128
 C;Accession: E75357
 Db 361 E 361
 RESULT 10
 H66891 exported serine proteinase (EC 3.4.21.-) [imported] - lactococcus lactis subsp. lactis C
 C;Species: Lactococcus lactis subsp. lactis
 C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C;Accession: H66891
 R;Bolotin, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrli, R.; Genome Res., 11, 731-753, 2001
 A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss H66891
 A;Reference number: A66625; MUID:21235186; PMID:1137471
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-408 <STOP>
 A;Experimental source: strain IL1403
 C;Genetics:
 C;Keywords: htrA
 C;Keywords: hydrolase; serine proteinase
 Query Match 19.8%; Score 133.5; DB 2; Length 408;
 Best local Similarity 31.1%; Pred. No. 0.0083; Matches 41; Conservative 20; Mismatches 48; Indels 23; Gaps 5;
 Db 13 GOGFAATPIGOAMATAQGQIRSGGGSPVTWGTGPTAFLGLGVVD-----NQMN----- 57
 C;Species: Xylella fastidiosa
 C;Accession: A82581
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A;Reference number: A82515; MUID:20365717
 A;Status: preliminary
 A;Accession: A82581
 A;Molecule type: DNA
 A;Residues: 1-514 <STOP>
 A;Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAFB5040.1; GSPDB:GN
 A;Experimental source: strain 9a5c
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, F.A.; Abencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.P.; Carrasco, D.M.; Carrer as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.; submitted to Genbank, June 2000
 A;Authors: Ferreira, V.C.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Freire, J.D.; Jungueira, M.L.; Kemper, B.L.; Kitajima, J.P.; Krieger, J.E.; Kuramame, E.E.; La chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C., F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, Rodrigues, V.; Rosa, A.J. de M.; de Rosa, R.V.; de Sa, R.G.; Santelli, R.V.; Silva, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva, M.; Tsukao, M.H.; Valiada, H.; van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Reference number: A89328
 A;Contents: annotation
 A;Genetics:
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Ventter, J.C.; Fraser, C.M.; Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896
 A;Accession: E75357
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-441 <STOP>
 A;Cross-references: GB:AE002017; GB:AE000513; NID:96459527; PIDN:AAFI1312.1; PID:9645953
 C;Genetics:
 C;Gene: DR1756
 C;Map position: 1
 C;Superfamily: proteinase hhb
 Query Match 19.4%; Score 131; DB 2; Length 441;
 C;Accession: F83550
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.; Dory, S.; Olson, M.V.; Nature 406, 959-964, 2000
 Query Match 19.4%; Score 131; DB 2; Length 441;
 Best Local Similarity 29.0%; Pred. No. 0.0015; Mismatches 47;保守性 24; 错配 43; 缺口 48; 填充 6;
 C;Accession: F83550
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.; Dory, S.; Olson, M.V.; Nature 406, 959-964, 2000
 RESULT 12
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 84
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAGSLQERVLATGEQPLRITVR-RGGKTRVEVRL 431
 Qy 6 NFQLSQGGG-----GFAATPIGOAMATAQGQINSGGGSPVTWGTGPTAFLGLGVVDNN---- 55
 Db 277 NTQILTGGQGSAGVGFAPINVKRLLPQLQAGK-----IST-GDVITAV 330
 Qy 56 -----GNGARVORVVGASAPAASLG-----IST-GDVITAV 84
 Db 331 PPQQLKAQGIPSSGALLQKQVYCGSPAAAGLRGGNNNGKLPSAQTSISPDGDLITAV 390
 Qy 85 DGAPINSATAADALNHHPGDVLSVTWOTKSGGTRGNVTL 126
 Db 391 NGQPLEDAG

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A;Reference number: A22950; MUID:20437337
A;Accession: F83550
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-474 <STO>
A;Experimental source: strain PA01
A;Genetics:
A;Gene: mucD; PA0766

Query Match 18.1%; Score 122.5; DB 2; Length 474;
Best Local Similarity 30.4%; Pred. No. 0.008; Matches 42; Conservatve 22; Mismatches 51; Indels 23; Gaps 5; Matches 42;

Qy 4 SDNRFQLSQQ-GGQFAPIDGQAMAIQOIRSGGGSPVHIGPTAFGLGVVDNN----- 55
Db 233 SQIIFTRSGFGMGLSFQAPIPDVIALNVQDOLKKAG--KVSRG---WLGWVIQEVNKILAES 286

Qy 56 -----NGNARVQRVYVGSAAPASLGSITDVTAVDAGPINSATAMADALNGHHPGDIVSV 110
Db 287 FGLDKPKPSALVAQIVEDQGPAAKGLQGDVILSUNGOSINESADLPHLVGNMKPGDKNL 346

Qy 111 ----TWORKSGGRTGNV 124
Db 347 DVIIRNGORKSLSMAVGSL 364

RESULT 14

T35666 probable integral membrane protein - *Streptomyces coelicolor*
C;Species: *Streptomyces coelicolor*
C;DB: 03-Dec-1999 #sequence_revision 03-Dec-1999
C;Accession: T35866
R;Subunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
Submitted to the EMBL Data Library, April 1999
A;Reference number: Z21591
A;Accession: T35866
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 1-203 <SAU>
A;Cross-references: EMLI:AU049727; PIDN:CARB41567.1; GSPDB:GN00070; SCOEDB:SC9B1.21
A;Experimental source: strain A3(2)
A;Genetics:
A;Gene: SCOEDB:SC9B1.21

Query Match 17.7%; Score 119.5; DB 2; Length 203;
Best Local Similarity 32.1%; Pred. No. 0.0057; Matches 36; Conservatve 15; Mismatches 56; Indels 5; Gaps 1; Matches 36;

Qy 12 GGCGCAPIPIGOVMAAGQIQRGGSPVHIGPT----AFLGQGVVDNGCARYVRVVG 66
Db 80 GGAGEQGRGTGATGRGAHPSAGPSAGRSSSPAPAGTGVVEANDKEPKARVGVHV 139

Qy 67 SARPAASLGISTGVITAVDGAPINSAATAMADALNGHHPGDIVSVTWTOKSGG 118
Db 140 PGPGYAAGLVGRGVLDLAVGVGTTRDSDLAHAVARAGPGKEVKLTVRHRSGG 191

RESULT 15

F87590 serine proteinase HtrA [imported] - *Caulobacter crescentus*
C;Species: *Caulobacter crescentus*
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87590
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.; Embley, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: F87590
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-530 <STO>
A;Cross-references: GB:AE005673; NID:913424352; PIDN:AAK24722.1; GSPDB:GN00148
C;Genetics: CC2158
C;Superfamily: Helicobacter serine proteinase
A;Gene: KS-GTPSESSIA

Query Match 17.4%; Score 117.5; DB 2; Length 530;
Best Local Similarity 31.6%; Pred. No. 0.023; Matches 42; Conservatve 18; Mismatches 50; Indels 23; Gaps 5; Matches 42;

Qy 13 GGGFAIPPIQAMAIQOIRSGGGSPVHIGPT--AF-----LGIVVVDNNNGCARYVOR 63
Db 283 GIGFAIPAEVAGVQKQLIENGKVVRGYIGVSIAMNAEMAELGMSDV---KGAIVAS 338

Qy 64 WVGSAAPASLGSITDVTAVDAGPINSATAMADALNGHHPGDIVSV-----WQT 114
Db 339 VPPGGPAKAGLPPDILVAVNGVKISDSSELTVSKARPGETIKVSIIRDGPRIVDV 398

Qy 115 KSGGTRTGIVNVL 127
Db 399 KS-GTPSESSIA 410

Search completed: August 6, 2002, 08:16:53
Job time: 253 sec

Scoring table:					
Searched:					
Total number of hits satisfying chosen parameters: 105224					
Minimum DB seq length: 0					
Maximum DB seq length: 200000000					
Post-processing: Minimum Match 0%					
Maximum Match 100%					
Listing first 45 summaries					
Database : SwissProt_40:*					
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.					
SUMMARIES					
Result #					
No.	Score	Query Match Length	DB ID	Description	
1	133.5	19.8	408	HTRA_LACLA	RESULT 1
2	125	18.5	413	1 HTRA_LACLA	HTRA_LACLA STANDARD; PRT: 408 AA.
3	118	17.5	413	1 DEGP_BRHAB	ID: HTRA_LACLA
4	113	16.7	513	1 DEGP_BRHAB	AC: 09la06; STANDARD; PRT: 408 AA.
5	106.5	15.8	355	1 DEGP_ECOLI	DT: 16-OCT-2001 (Rel. 40, Created)
6	105	15.6	455	1 DEGP_SALTY	DT: 16-OCT-2001 (Rel. 40, Last sequence update)
7	102.5	15.2	474	1 DEGP_ECOLI	DT: 01-MAR-2002 (Rel. 41, Last annotation update)
8	101.5	15.0	503	1 DEGP_BART	DE: Serine protease do like htra (EC 3.4.21.-) (HTRAL).
9	97	14.4	497	1 DEGP_CHLTER	GN: HTRA OR IL2136.
10	94.5	14.0	497	1 DEGP_CHLMD	OS: Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
11	94.5	14.0	957	1 Y228_MYCTU	OC: Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
12	94.5	14.0	466	1 HTRA_HAELN	OC: Lactobacillus.
13	92.5	13.7	448	1 DEGP_ARATH	OX: NCBI_TAXID=1360;
14	92	13.6	478	1 DEGP_BUGAI	RN: [1] "HTRA is the unique surface housekeeping protease in Lactococcus lactis and is required for natural protein processing.", Ruitenberg, Mol. Microbiol. 35:1042-1051(2000).
15	92	13.6	905	1 ZO3_MOUSE	RN: [2] SEQUENCE FROM N.A., AND CHARACTERIZATION.
16	89.5	13.3	905	1 ERS_RAT	RC: STRAIN=IL1403;
17	87	12.9	864	1 P115_CHICK	RC: MEDLINE=21235186; PubMed=11337471;
18	86	12.7	762	1 OMPA_RICRI	RA: Bolotin A., Winzer P., Mager S., Jaiillon O., Malarme K., Weissenbach J., Ehrlich S.D., Sorokin A.;
19	86	12.7	2349	1 P15921	RA: "The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403";
20	85	12.6	515	1 Y100_MYCNU	RA: Genome Res. 11:731-755 (2001).
21	85	12.6	2021	1 OMPA_RICCN	CC: FUNCTION: DEGRADES ABNORMAL EXPORTED PROTEINS. NEEDED FOR THE PRO-PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED PROTEINS.
22	83	12.3	488	1 DEGP_CHLPN	CC: SUBCELLULAR LOCATION: Membrane-bound (Probable).
23	83	12.3	510	1 MURF_MICTU	CC: -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE DEGP/DEGS FAMILY.
24	82	12.1	389	1 HMUX_DROME	CC: -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
25	81.5	12.1	219	1 VG36_BPT2	CC: This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
26	81.5	12.1	619	1 ELF1_HUMAN	CC: EMBL: AF115705; AAC006442; AAU022441; -.
27	81.5	12.1	714	1 CDG1_PAEMA	DR: EMBL: AF115705; AAC006442; AAU022441; -.
28	81	12.0	354	1 ODP_ZYMMO	DR: InterPro: IPRO01254; Protease2C.
29	81	12.0	430	1 TPSN_CHICK	DR: InterPro: IPRO01254; Trypsin.
30	80	11.9	382	1 YN28_MYCTU	DR: Pfam: PF00089; Trypsin_1.
31	79.5	11.8	292	1 PPNK_ECO57	DR: Pfam: PF00089; Trypsin_1.
32	79.5	11.8	292	1 PPNK_ECOLI	DR: SMART: SM00228; PDZ; 1.
33	79.5	11.8	1286	1 AIDA_ECOLI	DR: SMART: SM00228; PDZ; 1.

RP SPROTENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9743617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 050952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurakawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
 RX MEDLINE=88105815; PubMed=3322223;
 RA Vogel R.F., Entian K.-D., Mecke D.;
 RT "Cloning and sequence of the mdh structural gene of *Escherichia coli*
 coding for malate dehydrogenase.";
 Arch. Microbiol. 149:36-42(1987).
 RN [7]
 RP IDENTIFICATION.
 RA Bazan J.F., Flitterick R.J.;
 RT "Bacterial and catalytic models of trypsin-like viral proteases.";
 RL Semin. Virol. 1:311-322(1990).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sb-sib.ch).
 CC
 DR EMBL; U15661; AAC39931; .
 DR EMBL; U22495; AAC40061; .
 DR EMBL; U18997; AAC58037; .
 DR EMBL; AE00402; AAC762671; .
 DR EMBL; AB00551; AAC58331; .
 DR EMBL; AP002564; BAB37531; .
 DR EMBL; M24777; .; NOT_ANNOTATED_CDS.
 DR MEROPS; S01_275; .
 DR EcoGene; EG11652; hboB.
 DR InterPro; IPR01478; PDZ.
 DR InterPro; IPR01478; Protease2C.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PRO0505; PDZ; 1.
 DR Pfam; PF00088; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS50106; PDZ; 1.
 RC Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 28
 RA CHAIN 29 355
 FT DOMAIN 281 325
 FT ACT_SITE 96 96
 FT ACT_SITE 126 126
 FT ACT_SITE 201 201
 FT CONFLICT 253 253
 FT CONFLICT 307 307
 SQ SEQUENCE 355 AA; 37581 MW; D091BAD65B8FE1CC CRC64;

Query Match 15.8%; Score 106.5; DB 1; Length 355;
 Best Local Similarity 31.2%; Pred. No. 0.092; Mismatches 5;
 Matches 44; Conservative 10; Indels 23; Gaps 5;

Db 219 SFDKSNDDGETPEPGIGFAIPFQOLATKIMDKLIRDGRVIRGYGIGR---EIAPLHAQGS 274

QY 6 NFOLSQGQ---GRIFP-----IGQAMAIQGQRSGGSPPWIGPAPFLG 49

Db 275 GT---DLOGIIVNNEVFRPDGRANAGIQVNQNDLISUDNPKPAISALEMDVQAEIRPGSVIP 332

QY 110 VTWQTKSGGTRGIVNVLAEQG 130

Db 333 WV-VMRDDKKQLTLQVTOEXP 352

RESULT 6

ID DBQ_ECOLI STANDARD; PRT; 455 AA.

AC P39099; DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DBQ precursor (EC 3.4.21.).

GN DEGQ OR H30A OR B3234.

OS Escherichia coli.

OC Bacteria; Protoplasts; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI-Taxid=562;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / WI110;
 RC Bass S., Gu O., Goddard A.;
 RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.

RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degQ and degS, *Escherichia coli* genes encoding
 homologs of the *degP* protease.";
 RL J. Bacteriol. 178:1146-1153(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=7426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -!- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

- ID DEGP-ECOLI STANDARD; PRN; 474 AA.
 AC P09376; P15724;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT Protease domain precursor (EC 3.4.21.-)
 DEGP OR HtrA OR PTD OR B0161 OR Z0173 OR ECS0165.
 OS Escherichia coli, and
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI-TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=89057448; PubMed=3057437;
 RA Lipinska B., Sharma S., Georgopoulos C.;
 RT "Sequence analysis and regulation of the htrA gene of Escherichia
 coli: a sigma 32-independent mechanism of heat-inducible
 transcription.";
 RL Nucleic Acids Res. 16:10053-10067(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=94261330; PubMed=8202364;
 RA Fujita N., Mori H., Yura T., Ishihama A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 2.4-1.1 min (110,917-193,643 bp) region.";
 RL Nucleic Acids Res. 22:1637-1639(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vieira J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12.";
 RN Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Fedderspiel N., Hyman R., Kalmen S., Komar C., Kurdi O.,
 RA Jashkarli D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimaranta E.T., Potamakis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509052;
 RX MEDLINE=21156231; PubMed=11258795;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7, and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE OF 1-50 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90233597; PubMed=2165018;
- RN [9]
 RP CHARACTERIZATION, AND SEQUENCE OF 27-39.
 RX MEDLINE=90202693; PubMed=2180903;
 RA Lipinska B., Zyllicz M., Georgopoulos C.;
 RT "The HtrA (DegP) protein, essential for Escherichia coli survival at
 high temperatures, is an endopeptidase.";
 RL J. Bacteriol. 172:1791-1797(1990).
 RN [10]
 RP IDENTITY OF HTRA AND PROTEASE DO.
 RX MEDLINE=9122240; PubMed=202286;
 RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
 RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
 RT "Protease Do is essential for survival of Escherichia coli at high
 temperatures: its identity with the htrA gene product.";
 RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
 CC -I- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
 CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGQ.
 CC -I- SUBUNIT: MULTIMERIC.
 CC -I- SUBCELLULAR LOCATION: Periplasmic.
 CC -I- INDUCTION: BY HEAT SHOCK.
 CC -I- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
 CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
 CC -I- SIMILARITY: BELONGS TO PEPTIDE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGQ/DEGS FAMILY.
 CC -I- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M36536; AAA23994.1;
 DR EMBL; X12457; CA30997.1;
 DR EMBL; D26562; CAB0280.1; -.
 DR EMBL; AE000125; AAC73372.1; -.
 DR EMBL; U70214; AAB0591.1; -.
 DR EMBL; AE005192; AAG54665.1; -.
 DR EMBL; AP002550; BAB33588.1; -.
 DR EMBL; M29955; AA23717.1; -.
 DR EMBL; M31772; AA23680.1; -.
 DR PIR; SD1899; SD1899.
 DR MEROPS; S01_273; -.
 DR PIR; B35993; B35993.
 DR SWISS-2DPAGE; P09376; COLI.
 DR Ecogene; EG310463; degP.
 DR Interpro; IPR001478; PDZ.
 DR Interpro; IPR001940; protease2C.
 DR Interpro; IPR01254; Trypsin.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PRO0834; PROTEASE2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSTB; PS5106; PDZ; 2.
 KW Hydrolase; Serine protease; Heat shock; Periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL_1 26 PROTEASE DO.
 FT CHAIN 27 474 PROTEASE DO.
 FT DOMAIN 280 371 PDZ 1.

FT DOMAIN 377 466 PDZ 2.
 FT ACT_SITE 131 131 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 161 161 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 236 236 CHARGE RELAY SYSTEM (POTENTIAL).
 FT CONFLICT 10 10 A -> R (IN REF. 1, 7 AND 8).
 FT CONFLICT 46 46 E -> Q (IN REF. 1).
 FT CONFLICT 192 192 A -> G (IN REF. 1).
 FT CONFLICT 467 474 STYIILMQ -> RHLPVNAVATLNPLKTGRGSPVNL (IN REF. 1).
 FT SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;

Query Match 15.2%; Score 102.5; DB 1; Length 474;
 Best Local Similarity 30.4%; Pred. No. 0.25; Mismatches 14; Indels 33; Gaps 3;

Matches 31; Conservative 14; Mismatches 24; Indels 33; Gaps 3;

Qy 13 GQGFALP-----IQQAMATAGQTRGGGSPVHIGPTAFLGLGVVDNNGN----- 57
 Db 263 GIGFRIPSNMKVKNLTQMEYQVKRG-----ELIGMTELNSELAKAMKV 308
 Qy 58 ---GARVORVYGSAPASLGIISTGDTTAVDGAPINSATAM 95
 Db 309 DAQRGAFVSVQVLPNSSAAKAGIKAGDVVITSLNKGPISSFAAL 350

RESULT 9
 DEGP_BARHE STANDARD; PRT; 503 AA.
 ID DEGP_BARHE
 AC P54925; 01-OCT-1996 (Rel. 34. Created)
 DT 01-OCT-1996 (Rel. 34. Last sequence update)
 DT 16-OCT-2001 (Rel. 40. Last annotation update)
 DE probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
 DE (Antigen HtrA).
 DEGFP OR HtrA.
 OS Bartonella henselae (Rochalimaea henselae).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bartonellaceae; Bartonella.
 OX NCBI_TAXID=38323;
 RN [1]
 RP SEQUENCE FROM N_A.
 RC STRAIN-HOUSTON-1;
 RX MEDLINE=94299828; PubMed=8027347;
 RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
 RA Gorla S., Hager C., Edwards K.;
 RT "detection of Rochalimaea henselae DNA in specimens from cat scratch
 disease patients by PCR.";
 RT RL J. Clin. Microbiol. 32:942-948(1994).
 CC --!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC --!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 DEGP/DECO/DECS FAMILY.
 CC --!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; 120127; AAA97430.1; -. MEROPS; S01.273; -. DR Interpro; IPR001478; PDZ. DR Interpro; IPR001254; Trypsin. DR Interpro; IPR001440; Protease2C. DR Interpro; IPR001440; Trypsin. PFam; PF00595; PDZ; 2. PRIMIS; PR0034; PROTEASE2C. DR SMART; SM00238; PDZ; 2. DR ROSITE; PS50106; PDZ; 2. KW Hydrolase; Serine protease; Periplasmic; Signal. FT SIGNAL; 1 18 POTENTIAL.

FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-LIKE.
 FT DOMAIN 285 357 PDZ 1.
 FT DOMAIN 419 466 PDZ 2.
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F4743282A9E CRC64;

Query Match 15.0%; Score 101.5; DB 1; Length 503;
 Best Local Similarity 26.5%; Pred. No. 0.32; Mismatches 52; Indels 41; Gaps 6;

Matches 39; Conservative 15; Mismatches 52; Indels 41; Gaps 6;

Qy 7 FQLSOGQG-FAIPIQGQAMALIAGOQIRSGGGSPVHIGPTAFLGLGVVDNNGNARVQY 64
 Db 266 FSPSGCNVGTAAPATANEWVQOL-----IEKGLVORGWLGVQIOPV 309

Qy 65 -----VG-----SAPASLGISTGDTTAVDGAPINSATAMADALANGHHP 104
 Db 310 TKEISDSIGLKKEAKGALITDILKGPAKAGIKAGDVVLSVNGEKINDVRDALKRIANMSP 369

Qy 105 GDVISV-TWOKRSGGTRTGNNTLAEGP 130
 Db 370 GETVTLGVW--KSGKEENIKVLDSP 394

RESULT 10
 DEGP_CHLTR STANDARD; PRT; 497 AA.
 ID DEGP_CHLTR
 AC P18584; 08/03/00; 01-NOV-1990 (Rel. 16. Created)
 DT 30-MAY-2000 (Rel. 39. Last sequence update)
 DT 16-OCT-2001 (Rel. 40. Last annotation update)
 DE Probable serine protease DO-like precursor (EC 3.4.21.-) (59 kDa
 DE immunogenic protein) (SK59).
 GN DEGP OR HtrA OR Ctr823.
 OX Chlamydia trachomatis.
 OX Bacteria; Chlamydiiales; Chlamydiateae; Chlamydia.
 OX NCBI_TAXID=813;
 RN [1]
 RP SEQUENCE FROM N_A.
 RC STRAIN-SEROVAR 12;
 RX MEDLINE=90337348; PubMed=2279836;
 RA Kahane S., Weinstein Y., Sarov I.;
 RT "Cloning, characterization and sequence of a novel 59-kDa protein of
 Chlamydia trachomatis.";
 RT Gene 90:61-67(1990).
 RL [2]
 RN SEQUENCE FROM N_A.
 RP STRAIN-SEROVAR 12;
 RC STRAIN=DJUW-3(CX);
 RX MEDLINE=9900809; PubMed=9844136;
 RA Stephens R.S., Kalmus C.J., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger T., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.";
 RT Science 282:754-759(1998).
 CC --!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 DEGP/DEQ/DECS FAMILY.
 CC --!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC --!- CAUTION: REF. 1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
 CC THE AUTHORS TRANSLATED THEIR PROTEIN TO THAT OF WHAT SEEMS TO BE THE REAL
 CC PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC
DR EMBL; AE001355; AAC68420.1; -.
DR EMBL; M3119; AA23115.1; -.
DR InterPro; IPR00478; PDZ.
DR InterPro; IPR001940; protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00555; PDZ; 2.
DR Prints; PRO0834; PROTEASE2C.
DR Prints; PRO0839; V8PROTEASE.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
KW Hydrolase; Serine Protease; Signal; Antigen; Complete proteome.
FT SIGNAL 1 16 PROBABLE SERINE PROTEASE DO-LIKE.
FT CHAIN 17 497 CATALYTIC.
FT DOMAIN 128 289 PDZ 1.
FT DOMAIN 290 381 PDZ 2.
FT DOMAIN 394 485 PDZ 1.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
FT SEQUENCE 497 AA; 53244 MW; 86A5E31BBB4A38BA CRC64;
Query Match 14.4%; Score 97; DB 1; Length 497;
Best Local Similarity 29.9%; Pred. No. 0.73; Mismatches 57; Indels 14; Gaps 3;
Matches 38; Conservative 18; MisMatches 57; Indels 14; Gaps 3;
QY 13 GOGFAIPQGQAMIAQGQTTRSGGGSPPTVHIGPTAFLGLGVVDN-----NGNGARVQR 63
Db 274 GIGFIPSIMAKRVIDQLISDQVTGFLGVT---LOPDISELATCYKLEVKYGAIVTD 329
QY 64 WVGSAPAASLGISTGDTIATAVDGAPINSATAMADALNGHHPGDVLSWTWQTKGGRTGN 123
Db 330 WVGSPAEKAGLRLQEDIVTIVANGKEVESSLALRNATSLMMPGTRVVLKIVREGKIEP 388
QY 124 VTHAEGP 130
Db 389 VTVOIIP 395

RESULT 11
DEP_CHAMU STANDARD; PRT; 497 AA.
ID Q91I97; AC P56377; DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR TCO210.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83560;
RN SEQUENCE FROM N.A.
RC STRAIN=MOPN / Niigg;
MEDLINE=20130255; PubMed=10684935;
Read T.D., Brunham R.C., Sien C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S., Linher K., Weidman J., Khoi H., Croven B., Bowman C., Dodson R., Gwyn M., Nelson W., Debay R., Kolonay J., McClarty G., Salzberg S.L., Elsen J., Fraser C.M., "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39." Nucleic Acids Res. 28,1397-1406(2000); ALSO KNOWN AS THE DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
DR EMBL; AE002289; AAC39082.1; -.
DR TIGR; TC0210; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR Prints; PRO0834; PROTEASE2C.
DR Prints; PRO0839; V8PROTEASE.
DR SMART; SM00228; PDZ; 2.
DR Pfam; PF00595; trypsin; 1.
DR Prints; PRO0834; PROTEASE2C.
DR Prints; PRO0839; V8PROTEASE.
DR PROSITE; PS50106; PDZ; 2.
KW Hydrolase; Serine Protease; Signal; Complete proteome.
FT SIGNAL 1 16 PROBABLE SERINE PROTEASE DO-LIKE.
FT CHAIN 17 497 CATALYTIC.
FT DOMAIN 128 289 PDZ 1.
FT DOMAIN 290 381 PDZ 2.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
FT SEQUENCE 497 AA; 53294 MW; B765F350ACC6BBF CRC64;
Query Match 14.4%; Score 97; DB 1; Length 497;
Best Local Similarity 30.4%; Pred. No. 1.2; Mismatches 43; Indels 13; Gaps 2;
Matches 31; Conservative 15; MisMatches 43; Indels 13; Gaps 2;
QY 13 GOGFAIPQGQAMIAQGQTTRSGGGSPPTVHIGPTAFLGLGVVDN-----NGNGARVQR 63
Db 274 GIGFIPSIMAKRVIDQLISDQVTGFLGVT---LOPDISELATCYKLEVKYGAIVTD 329
QY 64 WVGSAPAASLGISTGDTIATAVDGAPINSATAMADALNGHHPG 105
Db 330 WVGSPAEKAGLRLQEDIVTIVANGKEVESSLALRNATSLMMPG 371

RESULT 12
Y278_MYCTU STANDARD; PRT; 957 AA.
ID Y278_MYCTU AC P56377; DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family Protein RV0278c Precursor.
GN RV0278C OR MT0931 OR MT0935_06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosh R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Egilmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsey T., Jacobs K., Krogh A., McLean J., Moulé S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Streeter S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.; Nature 393:537-544 (1998).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / oshkosh;

OX NCBI_TaxID=3702;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=CV; COLUMBIA;
RC Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2] SEQUENCE OF 91-113 AND 345-361.
RP
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RL Submitted (OCT-2000) to the SWISS-PROT data bank.
CC I- FUNCTION: PROBABLE SERINE PROTEASE.
CC -I- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID LUMEN (PROBABLE).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE DEGP/DEGO/DEGS FAMILY.
CC -I- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AB024023; BAA8101.1; - .
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; protease2C.
DR Pfam; PF00095; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PROSITE; PS50106; PDZ; 1.
DR KW hydrolase; serine protease; trans peptidase; Chloroplast; Thylakoid.
FT TRANSIT? 1 ?
FT CHAIN 91 448
FT DOMAIN 152 333
FT DOMAIN 335 433
FT ACT SITE 171 214
FT ACT SITE 292 292
FT SEQUENCE 448 AA; 47492 MW; A986FC1387670AF CRC64;

Query Match 13 6%; Score 92; DB 1; Length 448;
Best Local Similarity 29.5%; Pred. No. 1.6; Matches 36; Conservative 13; Mismatches 55; Indels 18; Gaps 4;

QY 10 SOGGQQFAPIPIQAMAIAGOIRGRRGGSPVHIGPTAFLGLGVDDN--NGARYQRVVG 66
Db 316 TSAVGFAPPSSTKIVPQLIQ--FSKVIRAGINIELADPVAQNOLNVRNGALMVLQVPG 373
QY 67 SAPAPASLG-----1STGDVITAVDGAINTSATAMADALGHHPGDIVSVWQTK 115
Db 374 KSLAEKAGLHPTSKRGFAGNIVLGLDIVAVDDKPKVNKAELMKILDYSGD--KVNLKIK 431
QY 116 SG 117
Db 432 RG 433

RESULT 15
DEGP_BUGAI STANDARD; PRT; 478 AA.

AC P57322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-Oct-001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).

GN DEGP OR BU228.
OS Buchnera aphidicola (subsp. *Acyrthosiphon pisum*) (*Acyrthosiphon pisum* symbiotic bacterium)
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RA Shigenobu S., Watanabe H., Hattori M., Sakai Y., Ishikawa H.;
RT "Genome Sequence of the endocellular bacterial symbiont of aphids Buchnera sp. Acs";
RL Nature 407:81-86(2000).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE DEGP/DEGO/DEGS FAMILY.
CC -I- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AP00118; BAB12943.1; - .
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR Pfam; PR00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR PROSITE; PS50106; PDZ; 1.
DR KW Hydrolase; Serine Protease; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 478
FT TRANSIT? 1 ?
FT DOMAIN 116 254
FT DOMAIN 281 372
FT DOMAIN 387 469
FT DOMAIN 401 500
FT ACT SITE 133 133
FT ACT SITE 163 163
FT ACT SITE 238 238
FT ACT SITE 238 238
FT SEQUENCE 478 AA; 52230 MW; 8688732CAC50629 CRC64;

Query Match 13 6%; Score 92; DB 1; Length 478;
Best Local Similarity 25.6%; Pred. No. 1.7; Matches 32; Conservative 23; Mismatches 50; Indels 20; Gaps 4;

QY 13 GOGEFAPIPIQAMAIAGOIRGRRGGSPVHIGPTAFLGLGVDDN--NINGNARVORY 64
Db 265 GIGFAIPCNWNKULTAQOMYQFG--QVRGEGLGMGMEINSLQAQIMKINSQKGAFVSRV 321
QY 65 VGGPAASLGSISTGDVITAVDGAINTSATAMADALGHHPGDIVSVWQTK 122
Db 322 LPNSSAFEAGIKAGDIISLRPKISSSSRLAEGS---LPVATKMELGVPREGI 375
QY 123 -NVTIL 126
Db 376 KNITV 380

Search completed: August 6, 2002, 08:17:54
Job time: 164 sec

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:12:10 ; Search time 16.84 Seconds
 (without alignments)
 191.460 Million cell updates/sec

Title: US-09-684-215a-4

Perfect score: 675

Sequence: 1 TAASDNFQLSOGQGQFAIPI.....QTKSGGRTGNVTLAECPPA 132

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

database : Issued_Patents_AA:^{*}

1: /cgn2_6/ptodata/2/iaa/5A_COMB_pep: *
 2: /cgn2_6/ptodata/2/iaa/5B_COMB_pep: *
 3: /cgn2_6/ptodata/2/iaa/5A_COMB_pep: *
 4: /cgn2_6/ptodata/2/iaa/6B_COMB_pep: *
 5: /cgn2_6/ptodata/2/iaa/PCTRIS_COMB_pep: *
 6: /cgn2_6/ptodata/2/iaa/backfilles1.pep: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	670	99.3	132	4	US-08-818-112-66
2	670	99.3	132	4	US-08-818-112-66
3	670	99.3	132	4	US-08-818-112-66
4	670	99.3	132	4	US-08-818-112-66
5	670	99.3	355	4	US-08-818-112-79
6	670	99.3	355	4	US-08-818-112-80
7	200.5	29.7	580	4	US-08-818-112-75
8	200.5	29.7	580	4	US-08-818-112-76
9	200.5	29.7	580	4	US-08-818-112-75
10	158.5	23.5	97	4	US-08-818-112-72
11	158.5	23.5	97	4	US-08-818-112-73
12	158.5	23.5	97	4	US-08-818-112-72
13	148	21.9	30	4	US-09-439-313-484
14	127.5	18.9	460	4	US-09-199-637A-132
15	108	475	1	US-08-278-091-6	
16	108	16.0	475	1	US-08-483-859-6
17	108	16.0	475	1	US-08-472-173-6
18	108	16.0	475	1	US-09-487-167-6
19	108	16.0	475	2	US-08-482-816-6
20	108	16.0	475	2	US-08-296-149-6
21	108	16.0	475	2	US-08-801-499-6
22	108	16.0	475	2	US-08-615-271-6
23	108	16.0	475	3	US-09-074-660-6
24	108	16.0	475	3	US-09-074-659-6
25	108	16.0	475	3	US-09-106-468-6
26	108	16.0	475	4	US-09-106-466A-6
27	108	16.0	475	4	US-09-106-467-6

RESULT 1
 DS-08-818-112-66

SEQUENCE 66, Application US/08818112
 ; Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Devin C.
 APPLICANT: Campos Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedula, Thomas S.
 APPLICANT: Twarzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,112
 FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 NAME: Makki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 66:

SEQUENCE CHARACTERISTICS:
 LENGTH: 132 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-818-112-66

Query Match 99.3%; Score 670; DB 4; length 132;
 Best Local Similarity 99.2%; Pred. No. 2.4e-62;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

APPLICANT: Vedrick, Thomas S.
 ADDRESS: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,112
 FILING DATE: 13-MAR-1997
 CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 622-4900
 INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:
 LENGTH: 355 amino acids
 STRANDEDNESS: single
 TOPOLOGY: linear

Query Match 99.3%; Score 670; DB 4; Length 355;
 Best Local Similarity 99.2%; Pred. No. 8.9e-62;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TAASDNFOUSOGGGGFAPIQGQAMATAGOIQRSGGSPTVHIGPTAFGLGIVUDNNNGAR 60
 Db 224 TAASDNFOUSOGGGGFAPIQGQAMATAGOIQRSGGSPTVHIGPTAFGLGIVUDNNNGAR 283

Qy 61 VORVVGSSAPAASLIGISNGDVITAVDGAPINSATAMADALNHHPGDVISVNWQTSGGTR 343
 Db 284 VORVVGSSAPAASLIGISNGDVITAVDGAPINSATAMADALNHHPGDVISVNWQTSGGTR 343

Qy 121 TGNVTLAEGPPA 132
 Db 344 TGNVTLAEGPPA 355

RESULT 5
 US-08-818-111-80
 Sequence 80, Application US/08818111
 Patent No. 6338852

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeky, Yasir A.W.
 APPLICANT: Dillon, Davin C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-APR-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392

RESULT 6
 US-09-056-556-79
 Sequence 79, Application US/09056556
 Patent No. 630456

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeky, Yasir A.W.
 APPLICANT: Dillon, Davin C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 NUMBER OF SEQUENCES: 148

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington

REFERENCE/DOCKET NUMBER: 210121.457
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 79:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-0-9-056-556-79

Query Match 99.3%; Score 670; DB 4; Length 355;
 Best Local Similarity 99.2%; Pred. No. 8.9e-62; Mismatches 0; Indels 0; Gaps 0;
 Matches 131; Conservative 0; MisMatches 1; Indels 0; Gaps 0;

Qy 1 TAASDNFQLSQGCGQFAPIPGQAMAIAGQIRSGGSPTVHIGPTAFLGLGVW-DNNG 57
 Db 224 TRASDNFQLSQGCGQFAPIPGQAMAIAGQIRSGGSPTVHIGPTAFLGLGVW-DNNG 57
 Qy 61 VORVVGSPAAASIGLSTGDTTAWDGAPINSATAMADALNHGGDVTWQKSGTR 120
 Db 284 VORVVGSPAAASIGLSTGDTTAWDGAPINSATAMADALNHGGDVTWQKSGTR 120
 Qy 121 TGNVTLAEGPPA 132
 Db 344 TGNVTLAEGPPA 355

RESULT 7
 US-08-818-112-75 ; Sequence 75, Application US/08818112
 Patent No. 6290969
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Twardzik, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 NUMBER OF SEQUENCES: 153
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patient In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,112
 FILING DATE: 13-MAR-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEX: (206) 682-6031
 INFORMATION FOR SSO ID NO: 76:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 580 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-818-111-76 ; Sequence 76, Application US/08818111
 Patent No. 6338852
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Vedwick, Thomas S.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 148
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

Query Match 29.7%; Score 200.5; DB 4; Length 580;
 Best Local Similarity 38.9%; Pred. No. 7.7e-13; Mismatches 50; Indels 9; Gaps 3;
 Matches 51; Conservative 21; MisMatches 50; Indels 9; Gaps 3;

Qy 1 TAASDNFQLSQGCGQFAPIPGQAMAIAGQIRSGGSPTVHIGPTAFLGLGVW-DNNG 57
 Qy 58 GARVORVVGSPAAASIGLSTGDTTAWDGAPINSATAMADALNHGGDVTWQKSGTR 117
 Db 507 GAKIVENVAGAAANAGVPKGVVVTKVDRPINSADALVAAVRSKAPGTVALTFQDPSG 566
 Db 453 TLGADSADAQSISGLGFAPIPVQDOKRIADELISTGKA-----SHASLGVOVTNDKDTP 506

RESULT 8
 US-08-818-111-76 ; Sequence 76, Application US/08818111
 Patent No. 6338852
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Vedwick, Thomas S.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 148
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patient In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,111
 FILING DATE: 13-MAR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEX: (206) 682-6031
 INFORMATION FOR SSO ID NO: 76:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 580 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-818-111-76

Query Match 29.7%; Score 200.5; DB 4; Length 580;
 Best Local Similarity 38.9%; Pred. No. 7.7e-13; Mismatches 50; Indels 9; Gaps 3;
 Matches 51; Conservative 21; MisMatches 50; Indels 9; Gaps 3;

Qy 1 TAASDNFQLSQGCGQFAPIPGQAMAIAGQIRSGGSPTVHIGPTAFLGLGVW-DNNG 57

RESULT 9
 US-09-056-556-75
 Sequence 75, Application US/09056556
 ; Patent No. 635056
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vediwick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 153
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 SYSTEM: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-APR-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REFERENCE/DOCKET NUMBER: 31,392
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-6031
 TELEFAX: (206) 682-4900
 INFORMATION FOR SEQ ID NO: 72:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 580 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-056-556-75

Query Match 29.7%; Score 200.5; DB 4; Length 580;
 Best local similarity 38.9%; Pred. No. 7.7e-13;
 Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

Query 1 TAASDNFOLSQG-GGFAPIPIGAMAIQOIRGGSPVHIGPTAFLGLGVW-DNNGN 57
 Db 453 TLGADSADAQSGSIGLGFALPVDQKRIADELISKKA-----SHASLGVOVTNDKDTP 506
 Qy 58 GARYQRVVGASAPASLGLISTGDVITAVDGPINSATAMADALNGHHRGDMVISVTWQTKSG 117
 Db 507 GAKTIVEWAGGAANAGVPGKVWVVTKVDRPINSADALVALVAVRSKPGATVALTFQDPMSG 566
 Qy 118 GTRTGNNVLAE 128
 Db 567 GSRTVQVQVTLGK 577

RESULT 10
 US-08-818-112-72
 Sequence 72, Application US/08818112
 ; Sequence 72, Application US/08818112

RESULT 11
 US-08-818-111-73
 Sequence 73, Application US/08818111
 ; Sequence 73, Application US/08818111
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vediwick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 148
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP

Patent No. 6290969
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vediwick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 153
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,112
 FILING DATE: 13-MAR-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 72:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 97 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-818-112-72

Query Match 23.5%; Score 158.5; DB 4; Length 97;
 Best local similarity 38.7%; Pred. No. 1.7e-09;
 Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

Query 39 WHQGTAFLGLGV--DNNNGARVORVVGASAPASLGLISTGDVITAVDGPINSATAM 95
 Db 2 ISTGCKASHASLGVOVTNDKDTPGAKIVEWAGGAANAGVPGKVWVTKVDRPINSADAL 61
 Qy 96 ADAALNGHHRGDDVISVTWQTKSGGPTGNVLAE 128
 Db 62 WAAYRSKARGATVALTFQDPMSGRTVQVTLGK 94

STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,111
 FILING DATE: 13-MAR-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Makki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C6
 TELEPHONE: (206) 522-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 73:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 97 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 818-111-73

Query Match 23.5% Score 158.5; DB 4; Length 97;
 Best Local Similarity 38.7%; Pred. No. 1.7e-09; Mismatches 38; Indels 3; Gaps 1;
 Matches 36; Conservative 16; ;
 QY 39 WHIGPAAFLGLGVV--DNNGNGARVQWVGASAPASLGISTGDTAVDGAPINSATM 95
 Db 2 ISTGKASHASLGQVQTNKDTPGAKIVEVAGAANAGVPKGWVTKVDRIPSADAL 61
 96 ADALNHHHPGDIVSVTQWTKGGTRTGNTLAE 128
 62 VAAVRSKAPGATVALTFQDPSSGSRTVQVTLSK 94

RESULT 13
 US-09-439-313-484
 Sequence 484, Application US/09439313
 Patent No. 6329505
 GENERAL INFORMATION:
 APPLICANT: Xu, Jiangchun
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Harlacher, Susan Louise
 APPLICANT: Jiang, Yuqui
 APPLICANT: Reed, Steven G.
 APPLICANT: Kalos, Michael
 APPLICANT: Fanger, Gary
 APPLICANT: Rettner, Mark
 APPLICANT: Soik, John
 APPLICANT: Day, Craig
 APPLICANT: Reed, John
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
 FILE REFERENCE: 210121.427C9
 CURRENT APPLICATION NUMBER: US/09/439,313
 CURRENT FILING DATE: 1999-11-12
 NUMBER OF SEQ ID NOS: 575
 SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ_ID NO: 484
 LENGTH: 30
 TYPE: PRT
 ORGANISM: Homo Sapien
 TREATMENT: US-09-439-313-484

Query Match 21.9% Score 148; DB 4; Length 30;
 Best Local Similarity 100.0%; Pred. No. 4.3e-09; Mismatches 0; Indels 0; Gaps 0;
 Matches 30; Conservative 0; ;
 QY 1 TAASDNFQLSOGQGQGFALPIQGQAMAIQOI 30
 Db 1 TRASDNFQLSOGQGQGFALPIQGQAMAIQOI 30

RESULT 14
 US-09-199-337A-132
 Sequence 132, Application US/09199637A
 Patent No. 6355411
 GENERAL INFORMATION:
 APPLICANT: Ausubel, Frederick
 APPLICANT: Goodman, Howard M.
 APPLICANT: Rahme, Laurence G.
 APPLICANT: Mahajan-Miklos, Shalina

APPLICANT: Tan, Man-Wah
 APPLICANT: Cao, Hui
 APPLICANT: Drenkard, Eliana
 APPLICANT: Tsongalis, John
 TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 FILE REFERENCE: 00786/363.002
 CURRENT APPLICATION NUMBER: US-09/199,637A
 CURRENT FILING DATE: 1998-11-25
 PRIORITY FILING DATE: 1997-11-25
 NUMBER OF SEQ ID NOS: 437
 SEQ ID NO: 132
 LENGTH: 460
 TYPE: PRP
 ORGANISM: Pseudomonas aeruginosa
 US-09-199-637A-132

Query Match
 Best Local Similarity 18.9%; Score 127.5; DB 4; Length 460;
 Matches 43; Conservative 21; Mismatches 51; Indels 23; Gaps 5;
 QY 4 SDNFQLSQC--GQFAPIQGQAMIAQGQRSGGSPTVHGPATRGLGVVDNNN-----
 Db 233 SQITPRSSGFMGLSFAPITDVALWADQLKKAG---KVSRG--WLGWVQIEVNKELAES 286
 QY 56 -----GNGARVQRVGQAPASASLIGTISTGQVITAVDGAPINSATAMADALNGHHPDSIVS 110
 Db 287 FGUDKPSGLVALQDVDEPAKGGIQLQVGIVLNSQINSESADLPHLVQNMKPGDDKINL 346
 QY 111 ---TQQRKSGGTRQGNV 124
 Db 347 DVIRNRGORKSLSMAVGNL 364

RESULT 15
 US-08-278-091-6
 Sequence 6 Application us/08278091
 Patent No. 5505139
 GENERAL INFORMATION:
 APPLICANT: LOOSMORE, Sheena M
 APPLICANT: YANG, Yan-Ping
 APPLICANT: CHONG, Pele
 APPLICANT: OOMEN, Raymond P.
 APPLICANT: KLEIN, Michel H.
 TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Activity
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.2.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-08/278.091
 FILING DATE: 21-JUL-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael T
 REGISTRATION NUMBER: 24,973
 TELECOMMUNICATION INFORMATION:
 REFERENCE/DOCKET NUMBER: 1038-371
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 6
 SEQUENCE CHARACTERISTICS:
 LENGTH: 475 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-278-091-6
 Query Match
 Best Local Similarity 28.9%; Pred. No. 0.0023; length 475;
 Matches 39; Conservative 18; Mismatches 44; Indels 34; Gaps 4;
 QY 13 GQGEAIP----IGQAMAIAGQIQRSGGSPTVHGPATRGLGVVDNNN-----
 Db 264 GIGFAIPSNNVKNLTSQMYEGQYRRG-----ELGIMTEHNSELAKAMVY 309
 QY 58 -----GARVQRVGQAPASASLIGTISTGQVITAVDGAPINSATAMADALNGHHPDSIVS 113
 Db 310 DAQRGAFYPSQMPNSAAKAGIKAGDVITSLSNGKPISSFAILRAQVGTMPVGSKISL-GL 368
 QY 114 TKGSGTRTGNTVLAE 128
 Db 369 LRECKAITVNLLEQ 383

Search completed: August 6, 2002, 08:16:20
 Job time: 250 sec

GenCore version 4.5
copyright (c) 1993 - 2000 Compugen Ltd.

Om protein - protein search, using sw model

Run on: August 6, 2002, 08:14:20 ; Search time 26.07 Seconds
(without alignments)

875.924 Million cell updates/sec

Title: US-09-684-215R-4
Perfect score: 675
Sequence: 1 TAASDPNQFQLSGGGCGFAIPL... QTKSGGTRTGNVTLAEGPPA 132

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBI_19:*

1: sp_archeal:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_micr:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description	RESULT ID	1	0.0175	PRELIMINARY:	PRT;	355 AA.	
1	675	100.0	355	16	007175	00175;						
2	474.5	70.3	361	2	Q50320	01-JUL-1997 (TREMBREL. 04, Last sequence update)						
3	459.5	68.1	354	16	Q9CCY9	01-DEC-2001 (TREMBREL. 19, Last annotation update)						
4	199.5	29.6	464	16	053896	DE HYPOTHETICAL 34.9 KDA PROTEIN.						
5	188.5	27.9	382	16	Q9CD67	PEPA OR RV0125 OR MTC418B.						
6	188.5	27.9	452	2	Q9Z5G6	Mycobacterium tuberculosis.						
7	161.5	23.9	542	2	Q9FBK9	Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.						
8	154.2	22.8	29	2	Q93J30	NCBI_TaxID=1773;						
9	139.0	20.6	394	16	P772780	[1]						
10	138.5	20.5	362	2	Q9S2R5	SEQUENCE FROM N.R.						
11	131.9	19.4	441	15	P0R01254	STRAIN:H37RV;						
12	128.5	19.0	514	16	Q9PBA3	MEDLINE-98295987; Pubmed=9634230;						
13	127.5	18.9	371	2	O31388	RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekla F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moulis S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Roberts J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;						
14	127.5	18.9	29	2	Q9AIS1	RT Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. ".						
15	122.5	18.1	474	16	P07155	Nature 393:537-544 (1998).						
16	17.9	513	16	Q9BCS8	R. -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.							
					DR EMBL: P26071; CABOB945.1; .	DR DR MEROPS: S01.UPC; .						
					DR Tuberquist; RV0125; .	DR DR InterPro: IPR00178; PDZ.						
					DR InterPro: IPR001254; TRYPSIN.	DR Pfnm; PR0059; PDZ; 1.						
					DR Pfam; PR00089; TRYPSIN; 1.	DR DR PROSITE; SM00228; PDZ; 1.						
					DR PROSITE; PS50106; PDZ; 1.	DR PROSITE; PS50240; TRYPSIN.DOM; 1.						

DR	PROSITE; PS00135; TRYPSIN_SER: UNKNOWN_1.	RESULT 3
KW	Complete protease; Hydrolase; Hypothetical protein; Serine protease.	09CCY9
SEQUENCE	355 AA; 34926 MW; 16C9E21K97BF192 CRC64;	Q9CCY9
SQ		ID Q9CCY9
Query Match	100 %; Score 675; DB 16; Length 355;	PRELIMINARY; PRT; 354 AA.
Best Local Similarity	100.0 %; Pred. No. 1. 5e-41; Mismatches 0; Indels 0; Gaps 0;	AC 09CCY9
Matches	132; Conservative 0; Mismatches 0;	DT DT 01-JUN-2001 (TREMBLrel. 17, Created)
QY	1 TAASINFOISQGGQFAIPQOMAIAQGQTAIRSGGSPTVHIGPFAFLGLGVVDNNNGAR 60	DT DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
Db	224 TAASDNPOLSQGGQFAIPQOMAIAQGQTAIRSGGSPTVHIGPFAFLGLGVVDNNNGAR 60	DE PROBABLE SECRETED SERINE PROTEASE.
QY	61 VORVVGSAAPASLGISTGVITAVDGAPINSATAMADALNGHHPGDVTSVTWOTKSGGR 120	GN GN M2659.
Db	284 VORVVGSAAPASLGISTGVITAVDGAPINSATAMADALNGHHPGDVTSVTWOTKSGGR 120	OS Mycobacterium leprae.
QY	121 TGNVTLAEGPPA 132	OC Bacteria; Firmicutes; Actinobacteria; Actinomycetidae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
Db	344 TGNVTLAEGPPA 355	OX NCBI_TAXID=1769;
RESULT	2	RN [1]
Q50320		RP SEQUENCE FROM N.A.
ID		RN STRAIN=TN;
Q50320	PRELIMINARY; PRT; 361 AA.	RX MEDLINE=21128732; PubMed=11234002;
ID		RA Cole S.T., Elgimelik K., Parkhill J., James K.D., Thomson N.R.,
Q50320		RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
DT	01-NOV-1996 (TREMBLrel. 01, created)	RA Munro K., Basham D., Brown D., Chillingworth T., Connor R.,
DT	01-NOV-1996 (TREMBLrel. 01, last sequence update)	RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hanafi N.,
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)	RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
DE	34KA PROBIN PRECURSOR.	RA Murphy L., Oliver K., Quail M.A., Rutherford K.M.,
OS	Mycobacterium paratuberculosis.	RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
OC	Bacteria; Firmicutes; Actinobacteria; Actinomycetidae; Mycobacteriaceae; Mycobacterium.	RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
OC		RA Barrell B.G.;
NCBI_TAXID=1770;		RT "Massive gene decay in the leprosy bacillus.";
RN	[1]	RL Nature 409:1007-1011 (2001).
RP	SEQUENCE FROM N.A.	CC -1 SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
RC	STRAN=JDB8/107;	DR EMBL; AU583926; CAQ32191.1; -.
RX	MEDLINE=95005419; PubMed=7921248;	DR MEROPS; S01_UPC; -.
AC	Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;	DR Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RT	"Identification and characterisation of a putative serine protease expressed in vivo by Mycobacterium avium subsp paratuberculosis";	DR InterPro; IPR001478; Protease2C.
RT	Microbiology 140:1977-1982 (1994).	DR InterPro; IPR001940; InterPro; IPR001256; Ser_proteas_v8.
EMBL	Z22092; CAA80638.1; -.	DR InterPro; IPR001254; Trypsin.
EMBL	S01_UPC; -.	DR Pfam; PF00595; PDZ; 1.
DR	InterPro; IPR001478; PDZ.	DR Pfam; PF00089; trypsin; 1.
DR	InterPro; IPR001254; trypsin.	DR PRINS; PRO0334; PROTEASE2C.
DR	Pfam; PF00595; PDZ; 1.	DR PRINS; PRO0839; v8PROBASE.
DR	Pfam; PF00089; trypsin; 1.	DR SMART; SM00228; PDZ; 1.
DR	SMART; SM00228; PDZ; 1.	DR PROSITE; PS00106; PDZ; 1.
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.	DR PROSITE; PS00240; TRYPSIN_DOM; 1.
KW	Complete protease; Hydrolase; Protease; Serine protease.	DR PROSITE; PS00240; TRYPSIN_DOM; 1.
SQ	SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;	DR PROSITE; PS00240; TRYPSIN_DOM; 1.
Query Match	68.1%; Score 459.5; DB 16; Length 354;	DR PROSITE; PS00240; TRYPSIN_DOM; 1.
Best Local Similarity	66.7%; Pred. No. 5.9e-26; Mismatches 88; Conservative 17; Indels 1; Gaps 1;	DR PROSITE; PS00240; TRYPSIN_DOM; 1.
Matches	26; Indels 1; Gaps 1;	DR PROSITE; PS00240; TRYPSIN_DOM; 1.
QY	1 TAASINFOISQGGQFAIPQOMAIAQGQTAIRSGGSPTVHIGPFAFLGLGVVDNNNGAR 60	QY 1 TAASINFOISQGGQFAIPQOMAIAQGQTAIRSGGSPTVHIGPFAFLGLGVVDNNNGAR 60
Db	224 TAASDNPOLSQGGQFAIPQOMAIAQGQTAIRSGGSPTVHIGPFAFLGLGVVDNNNGAR 60	Db 283 VARVATGFMAGISVGDIITSVDGVPISEATMTNVLPVPHRGDTAVWHFRVADGGR 289
QY	121 TGNVTLAEGPPA 132	QY 121 TGNVTLAEGPPA 132
Db	343 TGNVTLAEGPPA 354	Db 343 TGNVTLAEGPPA 354
RESULT	4	
QY	1 TAASINFOISQGGQFAIPQOMAIAQGQTAIRSGGSPTVHIGPFAFLGLGVVDNNNGAR 60	QY 61 VORVVGSAAPASLGISTGVITAVDGAPINSATAMADALNGHHPGDVTSVTWOTKSGGR 120
Db	231 TAATDSYKMS-GGCGFAIPQIGRAMAVANGNRSAGSNVTWIGPFAFLGLGVVDNNNGAR 289	Db 283 VARVATGFMAGISVGDIITSVDGVPISEATMTNVLPVPHRGDTAVWHFRVADGGR 289
QY	61 VORVVGSAAPASLGISTGVITAVDGAPINSATAMADALNGHHPGDVTSVTWOTKSGGR 120	QY 61 VORVVGSAAPASLGISTGVITAVDGAPINSATAMADALNGHHPGDVTSVTWOTKSGGR 120
Db	290 VORVVGSAAPASLGISTGVITAVDGAPINSATAMADALNGHHPGDVTSVTWOTKSGGR 349	Db 343 TGNVTLAEGPPA 354
QY	121 TGNVTLAEGPPA 132	QY 121 TGNVTLAEGPPA 132
Db	350 TGNVTLAEGPPA 361	Db 350 TGNVTLAEGPPA 361

OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TAXID=1773;
 RN [1] SEQUENCE FROM N.A.
 RP SPRAINH37R7;
 RX MEDLINE=98395987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglemeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., Maclean J., Moule S., Murphy L.,
 RA Oliver S., Quail M.A., Rajandream M.A., Rogers J.,
 RA Ritter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL021999; CAA17582.1; -.
 DR MEROPS; S01.UPC; -.
 DR InterPro; IPR01478; PDZ.
 DR InterPro; IPR01478; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 464 AA; 46452 MW; AE93B FCC53 EEC8F CRC64;

Query Match 29.6%; Score 199.5; DB 16; Length 464;
 Best Local Similarity 38.9%; Pred. No. 5.4e-07;
 Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TAASDNFQLSQG-GQGPAIPIGOAMAIAGOQIRSGGSPTVHGPATPLGLGVV-DNNGN 57
 DB 337 TLGADSAQDAGSISIGLFAIPVQDKRIADELISTG---SHASLGVOVNDKDTL 390

QY 58 GARVORVGAPSAPSLGIVSTGIVTAVDAGPNSATAMADALNHHDIVDVSIWTQWSG 117
 DB 391 GAKIVEVVAGGAANAGAAGVPKGVVTKVDRPINSADALVAAVRSKAAGTAVLTQDQPSG 450

QY 118 GTRGVNTLAE 128
 DB 451 GSTIVQWTLGK 461

RESULT 5
 Q9CD67 PRELIMINARY; PRT; 382 AA.
 AC Q9CD67;
 DT 01-JUN-2001 (TREMBrel. 17, Created)
 DT 01-JUN-2001 (TREMBrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)
 DE POSSIBLE SECRETED SERINE PROTEASE.
 DE ML0176.

OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TAXID=1769;
 RN [1] SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA Harris D., Taylor K.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
 RN [2] SEQUENCE FROM N.A.
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Submitter (FEB-1998) to the EMBL/GenBank/DDBJ databases.
 RN [3] SEQUENCE FROM N.A.
 RX MEDLINE=31128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mundall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmons M., Steflton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL583917; CAC29684.1; -.
 DR MEROPS; S01.UPC; -.
 DR Leptoma; ML0176; -.
 DR InterPro; IPR01478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001234; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00334; PROTEASE2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 382 AA; 37084 MW; 3D8DDDB8AE32A80D CRC64;

Query Match 27.9%; Score 188.5; DB 16; Length 382;
 Best Local Similarity 39.8%; Pred. No. 2.7e-06;
 Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

QY 2 AASINFQLSQGGGFAIPQAMAIAGOQIRSGGSPTVHGPATPLGLGVVWDNNGN-GAR 60
 DB 258 ADSGDAQDGSISIGLFAIPVQDKRIADELISTG---KATH---ASLGVOVATKGTPGAK 311

QY 61 VORVGSAPAASLGSISTGIVTAVDAGPNSATAMADALNHHDIVDVSIWTQWSG 120
 DB 312 VMDVYVAGGAANAAVAPKGVLTKVDRILLSSADALVAAVRSKAAGTAVLTQDQSGSR 371

QY 121 TGNYTLAE 128
 DB 372 TWQVTLGK 379

RESULT 6
 Q9Z5G6 PRELIMINARY; PRT; 452 AA.
 AC Q9Z5G6;
 DT 01-MAY-1999 (TREMBrel. 10, Created)
 DT 01-MAY-1999 (TREMBrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)
 DE PUTATIVE SERINE PROTEASE.
 GN MLC373_28.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TAXID=1769;
 RN [1] SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA Harris D., Taylor K.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
 RN [2] SEQUENCE FROM N.A.
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Submitter (FEB-1998) to the EMBL/GenBank/DDBJ databases.
 RN [3] SEQUENCE FROM N.A.
 RX MEDLINE=31128732; PubMed=8446027;
 RA Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
 RA "Use of an ordered cosmid library to deduce the genomic organization
 RT of Mycobacterium leprae";
 RL Mol. Microbiol. 7:197-205 (1993).
 DR EMBL; AL035500; CAB36690.1; -.
 DR MEROPS; S01.UPC; -.
 DR InterPro; IPR01478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001234; Trypsin.
 DR Pfam; PF00595; PDZ; 1.

Query Match 19.4%; Score 131; DB 16; Length 441;
Best Local Similarity 29.0%; Pred. No. 0 047; DB 16;
Matches 47; Conservative 24; Mismatches 43; Indels 48; Gaps 6;
QY 6 NFOQLSQQGQ-----GFATPIQAMAIAGOTRGSSGSPVHIGPTAFLGLGVVDNN---- 55
Db 277 NTQIITGGAGQSAGVYQFAIPNTVWPKLPOQAKG-----GVVSPSLGVFSDSLSSL 330
QY 56 -----GNGARVQRVQVGASAPASLG-----IST-GDVITAV 84
Db 331 PPOQOKKAAGLPLSSGALLQKQVPGSPSRAAGLRRGGNNGKLSLPSAQGSTSSISDGDLITAV 390
QY 85 DGAPNTSATAMADALINGHHGDVIVSVTWTQTKSGGTRITGNVT 126
Db 391 NGQPIEDAGSLOEAQVLAITGEQPLRUTVR-RGGKTREVEVTL 431

RESULT 12
Q9PRA3
ID Q9PRA3
AC Q9PRA3;
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE PERIPLASMIC PROTEASE.
GN XF2241.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=945C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acecinio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonacorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Boulo M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Hodeliel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieder J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite F.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
RA Marques M.V., Martins E.A.L., Martins M.M.F., Mattukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhanh A. Jr., Nobreza M.A., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V. de Rosa A.J.M.,
RA de Rosa V.E. Jr., da Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

Query Match 19.0%; Score 128.5; DB 16; Length 514;
Best Local Similarity 38.7%; Pred. No. 0 084; DB 16;
Matches 43; Conservative 11; Mismatches 46; Indels 11; Gaps 4;
QY 4 SDNFQLSQQGQ-----GFATPIQAMAIAGOTRGSSGSPVHIGPTAFLGLGVVDNN 54
Db 262 SOFSAASGGGMGISFAIPNLAINAQAQIRKTGKQRSRMGIVGPIDAKAOGQLGLPD 321
QY 55 NGNGARVQRVQVGASAPASLG-----IST-GDVITAV 105
Db 322 --RGALVNNIPPHSPAAKAGIEVGDBVIRSNGKVISSFSIDLPLIGMMPPG 370

RESULT 13
Q31388
ID Q31388
AC 031388;
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE DEGP PROTEIN.
GN DEGP.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium japonicum; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9611461; PubMed=9446679;
RA Narberhaus F., Weighofer W., Fischer H.M., Hennecke H.;
RT "Identification of the Bradyrhizobium japonicum degp gene as part of
an operon containing small heat shock protein genes";
RL Arch. Microbiol. 169:89-97(1998).
DR EMBL: Y13616; CAA73938.1; -.
DR ISSN: P31016; IBFE.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMURF: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR KW Complete proteome; Hydrolase; protease; Serine protease.
SQ SEQUENCE 441 AA; 45145 MW; D9E475AD081E0190 CRC64;

Query Match 18.9%; Score 127.5; DB 2; Length 371;
Best Local Similarity 28.9%; Pred. No. 0 059;

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terrenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zattz M., Medanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
RL Nature 406:151-159(2000).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AE002017; AACF1312.1; -.
DR MEROPS: SOLB.274; -.
DR TIGR: DR1756; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR PRINTS: PR00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMURF: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR KW Complete proteome; Hydrolase; Serine protease.
DR PRINTS: PR00839; VBPROTEASE.
DR SMURF: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR KW Complete proteome; Hydrolase; Serine protease.
SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82B4C CRC64;

Page 8

QY	56	-----GNGARVQRVVGSAAPASALGI	TGWDVTTAVDGAPI	SATAMADALINHNGHPGDIVSY	110
Db	287	FGLDKPSGAJVAQVLVEDGPAAKG	GQDVYDYLISLNQSQIN	ESADIPHLVGNMKGPKGDINKL	346
QY	111	---TWOTKSGGTRGVN	124		
Db	347	DVIRNCORKSLSMAVGSL	364		

Search completed: August 6, 2002, 08:17:32
Job time: 192 sec